Massachusetts Institute of Technology
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2011)

Problem Set 11 Solutions

1. Check book solutions on Stellar.

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

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

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

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)

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

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

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

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

Now we can find the conditional expectation estimator:

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

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

3. (a) Using the total probability theorem, we have

$$p_{T_1}(t) = \int_0^1 p_{T_1|Q}(t,q) f_Q(q) dq = \int_0^1 (1-q)^{t-1} q dq = \frac{1}{(t+1)t} \qquad \text{for } t = 1, 2, \ldots$$

(b) The least squares estimate coincides with the conditional expectation of Q given T_1, which is derived as

$$\begin{aligned} \mathbf{E}[Q \mid T_1 = t] &= \int_0^1 p_{Q|T_1}(q \mid t) q dq \\ &= \int_0^1 \frac{p_{T_1|Q}(t \mid q) f_Q(q)}{p_{T_1}(t)} q dq \\ &= \int_0^1 t(t+1) q (1-q)^{t-1} q dq \\ &= \int_0^1 t(t+1) q^2 (1-q)^{t-1} dq \\ &= t(t+1) \frac{2(t-1)!}{(t+2)!} \\ &= \frac{2}{t+2} \end{aligned}$$

(c) We write the posterior probability distribution of Q given $T_1 = t_1, \ldots, T_k = t_k$

$$\begin{aligned} f_{Q|T_1,\ldots,T_k}(q \mid t_1, \ldots, t_k) &= \frac{f_Q(q) \prod_i^k P_{T_i}(T_i = t_i \mid Q = q)}{\int_0^1 f_Q(q) \prod_i^k P_{T_i}(T_i = t_i \mid Q = q) dq} \\ &= \frac{q^k (1-q)^{\sum_i^k t_i - k}}{c} \\ &= \frac{1}{c} q^k (1-q)^{\sum_i^k t_i - k}, \end{aligned}$$

where the denominator integrates out q so it could be viewed as a constant scalar c. To maximize the above probability we set its derivative with respect to q to zero

$$kq^{k-1}(1-q)^{\sum_i^k t_i - k} - (\sum_i^k t_i - k)q^k(1-q)^{\sum_i^k t_i - k - 1} = 0,$$

or equivalently

$$k(1-q) - (\sum_i^k t_i - k)q = 0,$$

which yields the MAP estimate

$$\hat{q} = \frac{k}{\sum_{i=1}^k t_i}.$$

For this part only assume q is sampled from the random variable Q which is now uniformly distributed over $[0.5, 1]$

(d) The LLSE of T_1 given T_2 is

$$\hat{T}_2 = \mathbf{E}[T_2] + \frac{\text{cov}(T_1, T_2)}{\text{var}(T_1)}(T_1 - \mathbf{E}[T_1]),$$

where the coefficients are

$$\mathbf{E}[T_1] = \mathbf{E}[T_2] = \int_{0.5}^{1} f_Q(q)\mathbf{E}[T|Q=q]dq = \int_{0.5}^{1} 2 * 1/q dq = 2\ln 2,$$

and from the law of total variance

$$\begin{aligned}
\text{var}(T_1) = \text{var}(T_2) &= \mathbf{E}\left[\text{var}(T_1 \mid Q)\right] + \text{var}\left[\mathbf{E}(T_1 \mid Q)\right] \\
&= \mathbf{E}\left[\frac{1-Q}{Q^2}\right] + \text{var}\left[\frac{1}{Q}\right] \\
&= \mathbf{E}[1/Q^2] - \mathbf{E}[1/Q] + \mathbf{E}[1/Q^2] - \mathbf{E}[1/Q]^2 \\
&= \int_{0.5}^{2} f_Q(q)\frac{1}{q^2}dq - \int_{0.5}^{2} f_Q(q)\frac{1}{q}dq + \int_{0.5}^{2} f_Q(q)\frac{1}{q^2}dq - \left(\int_{0.5}^{2} f_Q(q)\frac{1}{q}dq\right)^2 \\
&= 2 - 2\ln 2 + 2 - (2\ln 2)^2 \\
&= 4 - 2\ln 2 - (2\ln 2)^2,
\end{aligned}$$

and their covariance

$$\begin{aligned}
\text{cov}(T_1, T_2) &= \mathbf{E}[T_1 T_2] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= \mathbf{E}\left[\mathbf{E}[T_1 T_2 \mid Q]\right] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= \mathbf{E}\left[\mathbf{E}[T_1 \mid Q]\mathbf{E}[T_2 \mid Q]\right] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= \mathbf{E}\left[1/Q^2]\right] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= 2 - 4(\ln 2)^2
\end{aligned}$$

Therefore we have derived the linear least squares estimator

$$\hat{T}_2 = 2\ln 2 + \frac{2 - 4(\ln 2)^2}{4 - 2\ln 2 - (2\ln 2)^2}(T_1 - 2\ln 2) \approx 1.543 + 0.113 T_1.$$

4. (a) Normalization of the distribution requires:

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

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

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

$$p_K(k;\theta) = \left(e^{-1/\theta}\right)^k \left(1 - e^{-1/\theta}\right), \qquad k = 0, 1, \ldots$$

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

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

and its variance is

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

(c) The joint probability distribution for the k_i is

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

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:

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

or

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

For a hot body, $\theta \gg 1$ and $\frac{1}{e^{1/\theta}-1} \approx \theta$, 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 for large enough n (in the limit) approaches 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

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