Massachusetts Institute of Technology

Department of Electrical Engineering & Computer Science

6.041/6.431: Probabilistic Systems Analysis

(Fall 2011)

Recitation 26 Solutions
December 13, 2011

1. (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

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

2. Let the true values of θ_0 and θ_1 be θ_0^* and θ_1^*, respectively. We have

$$\hat{\Theta}_1 = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2}, \qquad \hat{\Theta}_0 = \overline{Y} - \hat{\Theta}_1\overline{x},$$

where $\overline{Y} = (\sum_{i=1}^{n} Y_i)/n$, and where we treat $x_1, \ldots, x_n$ as constant. Denoting $\overline{W} = (\sum_{i=1}^{n} W_i)/n$, we have

$$Y_i = \theta_0^* + \theta_1^* x_i + W_i, \qquad \overline{Y} = \theta_0^* + \theta_1^*\overline{x} + \overline{W},$$

and

$$Y_i - \overline{Y} = \theta_1^*(x_i - \overline{x}) + (W_i - \overline{W}).$$

Thus,

$$\hat{\Theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(\theta_1^*(x_i - \overline{x}) + W_i - \overline{W})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \theta_1^* + \frac{\sum_{i=1}^{n}(x_i - \overline{x})(W_i - \overline{W})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$= \theta_1^* + \frac{\sum_{i=1}^{n}(x_i - \overline{x})W_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

where we have used the fact $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$. Since $\mathbf{E}[W_i] = 0$, it follows that

$$\mathbf{E}[\hat{\Theta}_1] = \theta_1^*.$$

Also

$$\hat{\Theta}_0 = \overline{Y} - \hat{\Theta}_1\overline{x} = \theta_0^* + \theta_1^*\overline{x} + \overline{W} - \hat{\Theta}_1\overline{x} = \theta_0^* + (\theta_1^* - \hat{\Theta}_1)\overline{x} + \overline{W},$$

and using the facts $\mathbf{E}[\hat{\Theta}_1] = \theta_1^*$ and $\mathbf{E}[\overline{W}] = 0$, we obtain

$$\mathbf{E}[\hat{\Theta}_0] = \theta_0^*.$$

Thus, the estimators $\hat{\Theta}_0$ and $\hat{\Theta}_1$ are unbiased.