PAP334 – Exercises 2 – Model answers

Problem 1

The height of the room is h, the length is l and the width is w. Their respective uncorrelated errors are σ_h, σ_l and σ_w. Because the errors are uncorrelated we can use the error propagation (see eqs. (1.57-58) in Cowan's Statistical Data Analysis textbook). It is assumed the room is rectangular and has no doors or windows.

The area of carpet needed A_c is (the carpet should cover the whole floor area):

$$A_c = l \times w.$$

The uncertainty on this latter can be evaluated given all random variables $x_i = l, w$:

$$\sigma_{A_c} = \sqrt{\sum_i \left(\frac{\partial A_c}{\partial x_i}\sigma_i\right)^2} = \left[\left(\frac{\partial A_c}{\partial l}\sigma_l\right)^2 + \left(\frac{\partial A_c}{\partial w}\sigma_w\right)^2\right]^{1/2} = \sqrt{(w\sigma_l)^2 + (l\sigma_w)^2} \quad \left(= A_c\sqrt{\left(\frac{\sigma_l}{l}\right)^2 + \left(\frac{\sigma_w}{w}\right)^2}\right)$$

The same procedure can be followed to evaluate the uncertainty on the area

$$A_w = 2\,(l \times h + w \times h)$$

of wallpaper needed to cover the room

$$\sigma_{A_w} = \sqrt{(2h\sigma_l)^2 + (2(l+w)\sigma_h)^2 + (2h\sigma_w)^2} = 2\sqrt{h^2\left(\sigma_l^2 + \sigma_w^2\right) + (l+w)^2\sigma_h^2}.$$

The correlation between two quantities A_c and A_w, functions of the three random variables quoted above, can be studied through the 2×2 covariance matrix

$$\mathscr{U} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} = \begin{pmatrix} \sigma_{A_c}^2 & U_{cw} \\ U_{wc} & \sigma_{A_w}^2 \end{pmatrix},$$

with

$$\begin{aligned} U_{cw} = U_{wc} &= \sum_i \left(\frac{\partial A_c}{\partial x_i}\right)\left(\frac{\partial A_w}{\partial x_i}\right)\sigma_i^2 = \frac{\partial A_c}{\partial w}\frac{\partial A_w}{\partial w}\sigma_w^2 + \frac{\partial A_c}{\partial l}\frac{\partial A_w}{\partial l}\sigma_l^2 + \frac{\partial A_c}{\partial h}\frac{\partial A_w}{\partial h}\sigma_h^2 \\ &= 2h\left(l\sigma_w^2 + w\sigma_h^2\right), \end{aligned}$$

which is positive all over the "physical" range of parameters, and only null'ed if $h = 0$ or all three variables are null'ed. Therefore, A_c and A_w are clearly correlated.

Finally, the correlation coefficient ρ_{cw} can be computed using

$$\begin{aligned} \rho_{cw} = \frac{U_{cw}}{\sigma_{A_c}\sigma_{A_w}} &= \frac{h(l\sigma_w^2 + w\sigma_l^2)}{\sqrt{\left[(l\sigma_w)^2 + (w\sigma_l)^2\right] \cdot \left[h^2\left(\sigma_l^2 + \sigma_w^2\right) + (l+w)^2\sigma_h^2\right]}} \\ &= \frac{h(l+w)}{\sqrt{(l^2+w^2)\cdot(2h^2 + (l+w)^2)}}, \text{ for } \sigma_w = \sigma_l = \sigma_h \equiv \sigma, \end{aligned}$$

hence is fully independent from the measurements uncertainties, and diverging for l, w and $h \to 0$.

This correlation coefficient is shown in Figure 1 for $l, w \in [2, 10]$, and $h = 2.4$ m. As previously mentioned, it can be seen that the maximal correlation is obtained for $h, l, w \to 0$.

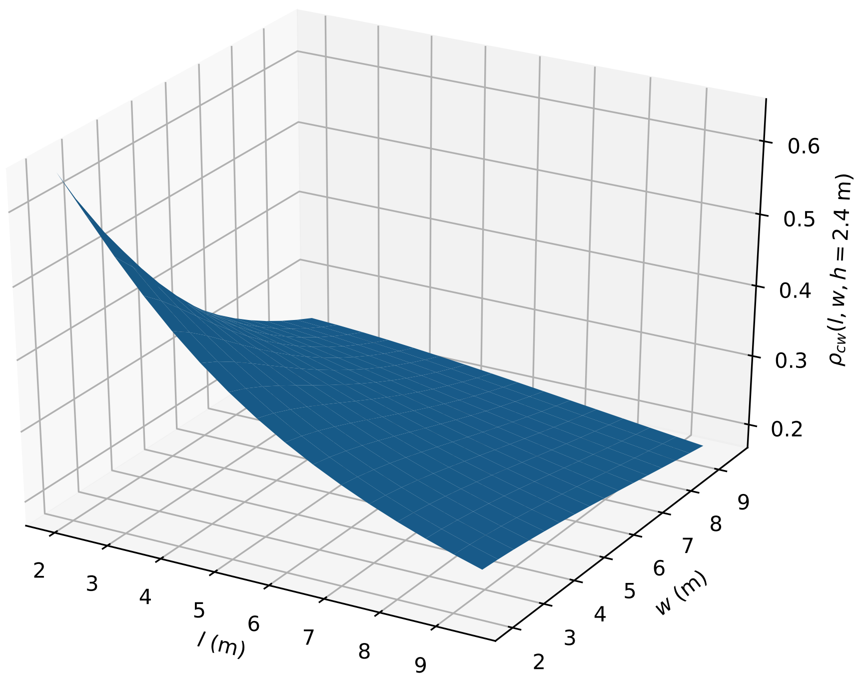

Figure 1: A visual representation of the analytic correlation coefficient ρ_{cw} calculated above for a small range of l and w, and for a fixed $h = 2.4$ m.

Problem 2

Northern lights are said to occur in Saariselkä with $P(A) = 1/2$ probability each night. They can be seen if they occur and the sky is clear enough, for which the probability is $P(W) = 3/5$. The time spent in Saariselkä is $N = 3$ nights.

(a) The probability to observe Northern lights every night during the visit can be computed knowing the probability to see them over any single night. As the Northern lights happening is uncorrelated with a clear sky to occur, this is expressed as:

$$P(\text{single night}) = P(A) \cdot P(W) = 0.3.$$

In general the probability for a combination of n occurences of a phenomenon with a probability p over a total of N nights can be expressed using the binomial (discrete) distribution:

$$f(n|p, N) = \frac{N!}{n!(N-n)!} p(1-p)^{N-n}.$$

However, for the simple particular cases studied here, one can simply use:

(i) $P(\text{every night}) = P(\text{single night})^N = 0.3^3 = 2.7\%$, and

(ii) $P(\text{no observation}) = P(\bar{A}) \cdot P(\bar{W}) + P(\bar{A}) \cdot P(W) + P(A) \cdot P(\bar{W}) = 70\%$. This latter result can also be obtained using the complement of $P(\text{single night})$, hence $P(\text{no observation}) = 1 - 0.3 = 0.7$.

(b) One may remember the analytical form of the Poisson (discrete), and Gaussian (continuous) probability density functions, on top of the binomial (discrete) distribution already introduced above:

$$f_{\text{Poisson}}(n|\nu) = \frac{\nu^n}{n!}e^{-\nu}$$
$$f_{\text{Gaussian}}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

For the earlier, the parameter ν corresponds both to the mean and variance of the distribution, and $n \in \mathbb{Z}^+$. The latter, with a mean μ and variance σ^2 is defined for all $x \in \mathbb{R}$.

(i) As shown in Figure 2 the Gaussian is relatively similar to the Poisson distribution, for $\mu = \sigma^2 = \nu = 7$. However, the Poisson distribution is slightly asymmetric with a shift of its maximum towards lower values of k than the Gaussian.

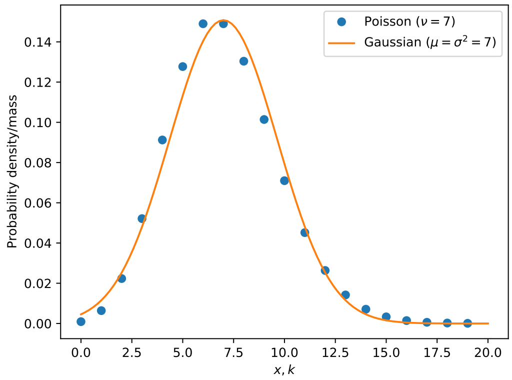

Figure 2: Comparison between a Gaussian (continuous) and a Poisson (discrete) distribution, with equivalent parameters ($\mu = \sigma^2 = \nu$).

(ii) The binomial with 14 trials and the Gaussian are in not in agreement, according to the leftmost part of Figure 3. According to the rightmost part of Figure 3 the binomial with 1200 trials and the Gaussian, however, are in relatively good agreement. The Gaussian thus looks like a sufficiently good approximation for the binomial distributions with large number of trials.

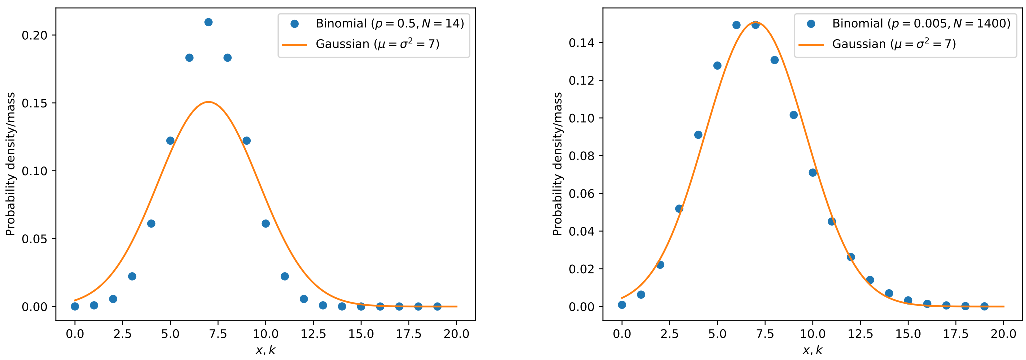

Figure 3: Comparison between a Gaussian (continuous) and two binomial (discrete) distributions: one with a high probability of occurence and a low trials multiplicity (left), and another with a low probability and high multiplicity (right).

(iii) The comparison between the Poisson and the binomial distributions in Figure 4 shows that the one with 1400 trials is very similar to the Poisson distribution, but the other one is not. This is exactly as expected; a binomial distribution is similar to the Poisson distribution when N is large, but p is small.

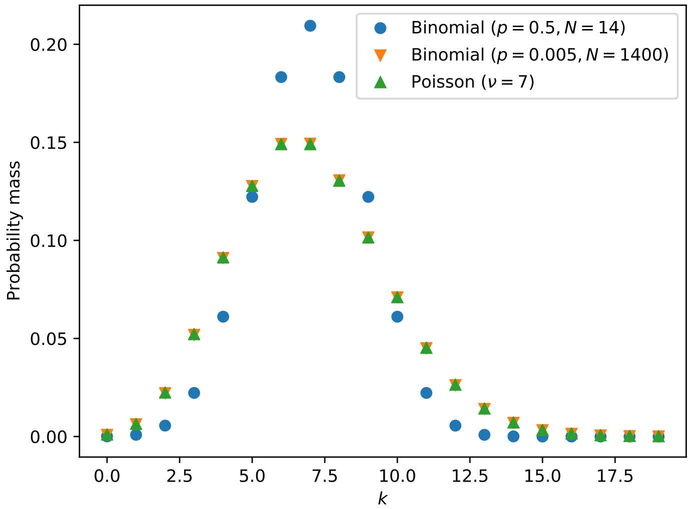

Figure 4: Comparison between a Poisson and two binomial distributions: one with a high probability of occurence and a low trials multiplicity, and another with a low probability and high multiplicity.

Code listing

Problem 1

```
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np

# initialise the plotting part

h = 2.4 # m
l = np.arange(2., 10., 0.5)
w = np.arange(2., 10., 0.5)
print(l,w)
l, w = np.meshgrid(l, w)

corr = h*(l+w)/np.sqrt((l**2+w**2)*(2*h**2+(l+w)**2))

fig = plt.figure()
ax = Axes3D(fig)
ax.set_xlabel('$l$ (m)')
ax.set_ylabel('$w$ (m)')
ax.set_zlabel('$\\rho_{cw}(l,w,h=2.4~{\\rm m})$')
ax.plot_surface(l, w, corr, rstride=1, cstride=1)
plt.show()
```

Problem 2

```
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats

k = np.arange(0., 20.)
x = np.linspace(0., 20., 101)
norm = stats.norm(loc=7., scale=np.sqrt(7.))

fig1 = plt.figure(1)
pois = stats.poisson(mu=7.)
plt.plot(k, pois.pmf(k), 'o', label='Poisson ($\\nu=7$)')
plt.plot(x, norm.pdf(x), label='Gaussian ($\mu=\sigma^2=7)$')
plt.xlabel('$x,k$')
plt.ylabel('Probability density/mass')
plt.legend(loc='best')
plt.show()

fig2 = plt.figure(2)
binom1 = stats.binom(p=0.5, n=14)
plt.plot(k, binom1.pmf(k), 'o', label='Binomial ($p=0.5, N=14$)')
plt.plot(x, norm.pdf(x), label='Gaussian ($\mu=\sigma^2=7)$')
plt.xlabel('$x,k$')
plt.ylabel('Probability density/mass')
plt.legend(loc='best')
plt.show()

fig3 = plt.figure(3)
binom2 = stats.binom(p=0.005, n=1400)
plt.plot(k, binom2.pmf(k), 'o', label='Binomial ($p=0.005, N=1400$)')
plt.plot(x, norm.pdf(x), label='Gaussian ($\mu=\sigma^2=7)$')
plt.xlabel('$x,k$')
plt.ylabel('Probability density/mass')
plt.legend(loc='best')
plt.show()

fig4 = plt.figure(4)
plt.plot(k, binom1.pmf(k), 'o', label='Binomial ($p=0.5, N=14$)')
plt.plot(k, binom2.pmf(k), 'v', label='Binomial ($p=0.005, N=1400$)')
plt.plot(k, pois.pmf(k), '^', label='Poisson ($\\nu=7$)')
plt.xlabel('$k$')
plt.ylabel('Probability mass')
plt.legend(loc='best')
plt.show()
```