Exercise 1 (Section 2.1)**.** Recall that the r-th central moment of a discrete random variable Y described by a probability mass function $f(y)$ is given by

$$\mathbb{E}(Y^r) = \sum_{y \in P} y^r f(y). \tag{1}$$

The r-th moment about the mean $\mu = \mathbb{E}(Y)$ of Y is

$$\mathbb{E}((Y-\mu)^r) = \sum_{y \in P} (y-\mu)^r f(y). \tag{2}$$

Let $Y_1, Y_2, \dots Y_n$ be discrete random variables described by a joint probability mass function $f(y_1, y_2, \dots, y_n)$. For any function $g(Y_1, Y_2, \dots, Y_n)$ its expectation value is

$$\mathbb{E}(g(Y_1, Y_2, \dots, Y_n)) = \sum_{y_1, y_2, \dots, y_n \in P} g(y_1, y_2, \dots, y_n) f(y_1, y_2, \dots, y_n). \tag{3}$$

Show that:

1. $\mathbb{E}(\alpha_0 + \alpha_1 Y_1^r + \alpha_2 Y_2^p) = \alpha_0 + \alpha_1 \mathbb{E}(Y_1^r) + \alpha_2 \mathbb{E}(Y_2^p)$, where $\alpha_0, \alpha_1, \alpha_2$ are constants. In other words, you need to show that $\mathbb{E}$ is a linear operator.
2. $\sigma^2 := \mathbb{E}((Y-\mu)^2) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$.
3. $\mathrm{Cov}(Y_1, Y_2) := \mathbb{E}((Y_1 - \mu_1)(Y_2 - \mu_2)) = \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2)$.
4. $\mathrm{Cov}(Y_1, Y_2) = 0$ if Y_1, Y_2 are independent random variables. Hint: use the property that the joint probability mass function factorizes as $f(y_1, y_2) = f_{Y_1}(y) f_{Y_2}(y)$ if Y_1 and Y_2 are independent random variables.

Solution.

1. Need to use (3) with $g(Y_1, Y_2) = \alpha_0 + \alpha_1 Y_1^r + \alpha_2 Y_2^p$ and then (1):

$$\begin{aligned}
\mathbb{E}(\alpha_0 + \alpha_1 Y_1^r + \alpha_2 Y_2^p) &= \sum_{y_1, y_2 \in P} (\alpha_0 + \alpha_1 y_1^r + \alpha_2 y_2^p) f(y_1, y_2) \\
&= \alpha_0 \underbrace{\sum_{y_1, y_2 \in P} f(y_1, y_2)}_{=1} + \alpha_1 \sum_{y_1, y_2 \in P} y_1^r f(y_1, y_2) + \alpha_2 \sum_{y_1, y_2 \in P} y_2^p f(y_1, y_2) \\
&= \alpha_0 + \alpha_1 \sum_{y_1 \in P} y_1^r f_{Y_1}(y_1) + \alpha_2 \sum_{y_2 \in P} y_2^p f_{Y_2}(y_2) \\
&= \alpha_0 + \alpha_1 \mathbb{E}(Y_1^r) + \alpha_2 \mathbb{E}(Y_2^p)
\end{aligned}$$

where

$$f_{Y_1}(y_1) = \sum_{y_2 \in P} f(y_1, y_2) \qquad f_{Y_2}(y_2) = \sum_{y_1 \in P} f(y_1, y_2)$$

are marginal mass functions of Y_1 and Y_2, respectively.

2. Need to use (1), (3) and $\mu := \mathbb{E}(Y)$

$$\begin{aligned}
\mathbb{E}((Y-\mu)^2) &= \sum_{y\in P}(y-\mu)^2 f(y) \\
&= \sum_{y\in P}(y^2 - 2y\mu + \mu^2) f(y) \\
&= \underbrace{\sum_{y\in P} y^2 f(y)}_{=\mathbb{E}(Y^2)} - 2\mu \underbrace{\sum_{y\in P} y f(y)}_{=\mathbb{E}(Y)} + \mu^2 \underbrace{\sum_{y\in P} f(y)}_{=1} \\
&= \mathbb{E}(Y^2) - 2\mu\cdot\mu + \mu^2 \cdot 1 \\
&= \mathbb{E}(Y^2) - \mu^2 \\
&= \mathbb{E}(Y^2) - \mu^2 = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2.
\end{aligned}$$

3. Need to use (1) and (3) once again:

$$\begin{aligned}
\mathrm{Cov}(Y_1, Y_2) &= \mathbb{E}((Y_1-\mu_1)(Y_2-\mu_2)) \\
&= \sum_{y_1, y_2\in P} (y_1-\mu_1)(y_2-\mu_2) f(y_1, y_2) \\
&= \sum_{y_1, y_2\in P} (y_1 y_2 - y_1\mu_2 - y_2\mu_1 + \mu_1\mu_2) f(y_1, y_2) \\
&= \sum_{y_1, y_2\in P} y_1 y_2 f(y_1, y_2) - \mu_2 \sum_{y_1, y_2\in P} y_1 f(y_1, y_2) \\
&\qquad - \mu_1 \sum_{y_1, y_2\in P} y_2 f(y_1, y_2) + \mu_1\mu_2 \sum_{y_1, y_2\in P} f(y_1, y_2) \\
&= \mathbb{E}(Y_1 Y_2) - \mu_2\,\mathbb{E}(Y_1) - \mu_1\,\mathbb{E}(Y_2) + \mu_1\mu_2 \\
&= \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_2)\,\mathbb{E}(Y_1) - \mathbb{E}(Y_1)\,\mathbb{E}(Y_2) + \mathbb{E}(Y_1)\,\mathbb{E}(Y_2) \\
&= \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\,\mathbb{E}(Y_2).
\end{aligned}$$

4. We need to show that $\mathbb{E}(Y_1 Y_2) = \mathbb{E}(Y_1)\,\mathbb{E}(Y_2)$ if $f(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$:

$$\mathbb{E}(Y_1 Y_2) = \sum_{y_1, y_2\in P} y_1 y_2 f(y_1, y_2) = \sum_{y_1\in P} y_1 f_{Y_1}(y_1) \sum_{y_2\in P} y_2 f_{Y_2}(y_2) = \mathbb{E}(Y_1)\,\mathbb{E}(Y_2).$$

Exercise 2 (Section 2.3)**.** Let Y be a Poisson random variable parameterised by a constant θ. Its probability mass function is $f(y) = \theta^y e^{-\theta}/y!$ where $y = 0, 1, 2, \ldots$

1. Show that $\mathbb{E}(Y) = \theta$. Hint: use Taylor series of the exponential function

$$e^\theta = \sum_{i=0} \frac{\theta^i}{i!} = 1 + \theta/1! + \theta^2/2! + \ldots$$

2. Find the maximum likelihood estimate of θ. Hint: consider a sample $y_1, y_2, \ldots y_n$ drawn from $Y_1, Y_2, \ldots, Y_n$ which are i.i.d. Poisson random variables parameterised by θ.
3. Minimize $S = \sum_{i=1}^n (y_i - \theta)^2$ to obtain the least squares estimate of θ.

Trivia: The Poisson distribution, for example, is appropriate for modeling the number of phone calls an office would receive during the noon hour, or the number of goals scored in each game in the World Cup, with θ being the expected value (average). You will study Poisson regression in the 3rd year "Linear Modelling" course.

Solution.

1. By definition,

$$\mathbb{E}(Y) = \sum_{y \in P} y f(y) = \sum_{y=0}^{\infty} \left(y \cdot \frac{\theta^y e^{-\theta}}{y!} \right)$$

Notice that $\left(y \cdot \theta^y e^{-\theta}/y!\right) = 0$ if $y = 0$ hence we can replace $\sum_{y=0}^{\infty}$ with $\sum_{y=1}^{\infty}$. We can also pull $e^{-\theta}$ outside the sum since it is independent of y. This gives

$$\sum_{y=0}^{\infty} \left(y \cdot \frac{\theta^y e^{-\theta}}{y!} \right) = e^{-\theta} \sum_{y=1}^{\infty} \frac{y \cdot \theta^y}{y!} = e^{-\theta} \sum_{y=1}^{\infty} \frac{\theta^y}{(y-1)!}$$

Next, write $\theta^y = \theta \cdot \theta^{y-1}$, so that

$$e^{-\theta} \sum_{y=1}^{\infty} \frac{\theta^y}{(y-1)!} = \theta e^{-\theta} \sum_{y=1}^{\infty} \frac{\theta^{y-1}}{(y-1)!} = \theta e^{-\theta} \sum_{y=0}^{\infty} \frac{\theta^y}{y!}$$

where in the last step we simply replaced y with $y + 1$. The last sum is nothing but e^θ, giving:

$$\mathbb{E}(Y_i) = \theta e^{-\theta} \sum_{y=0}^{\infty} \frac{\theta^y}{y!} = \theta e^{-\theta} e^{\theta} = \theta.$$

2. To find the maximum likelihood estimate (m.l.e.) of θ we need to find the likelihood $L(\theta)$, then compute its log-likelihood, $l(\theta) = \log(L(\theta))$, and find a value of θ that maximizes $l(\theta)$ for a given sample.

By definition, the likelihood $L(\theta)$ for a sample $y_1, \dots, y_n$ from $Y_1, \dots, Y_n$ is given by

$$L(\theta) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} = \underbrace{\prod_{i=1}^{n} e^{-\theta}}_{=e^{-n\theta}} \cdot \prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i!} = e^{-n\theta} \prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i!}$$

Here we used the product factorisation property, $\prod_i a_i b_i = \prod_i a_i \cdot \prod_i b_i$.

The log-likelihood is

$$\begin{aligned} l(\theta) = \log(L(\theta)) &= \log e^{-n\theta} + \sum_{i=1}^{n} \log(\theta^{y_i}/y_i!) \\ &= -n\theta + \sum_{i=1}^{n} y_i \log\theta - \sum_{i=1}^{n} \log y_i! \end{aligned}$$

Here we used

$$\log ab = \log a + \log b, \quad \log a/b = \log a - \log b, \quad \log a^b = b \log a, \quad \log e^a = a$$

Next, we compute the derivative of log-likelihood with respect to θ:

$$\frac{\partial l(\theta)}{\partial \theta} = -n + \sum_{i=1}^{n} \frac{y_i}{\theta} - 0 = -n + \frac{n\bar{y}}{\theta}.$$

Requiring $\frac{\partial l(\theta)}{\partial \theta}|_{\theta=\hat{\theta}} = 0$ we find that the m.l.e. $\hat{\theta}$ of θ:

$$-n + \frac{n\bar{y}}{\hat{\theta}} = 0 \quad \Longrightarrow \quad \hat{\theta} = \bar{y}$$

The hat $\hat{}$ is used to distinguish the estimated value $\hat{\theta}$ from the true unknown θ.

We also need to conduct the second derivative test to ensure that $\theta = \hat{\theta} = \bar{y}$ is a local maximum of $l(\theta)$:

$$\left.\frac{\partial^2 l(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial}{\partial \theta}\left(-n + \frac{n\bar{y}}{\theta}\right)\right|_{\theta=\hat{\theta}} = \left.-\frac{n\bar{y}}{\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{n}{\bar{y}} < 0$$

since $\bar{y} > 0$. Thus $\hat{\theta} = \bar{y}$ is indeed the m.l.e. of θ.

3. To minimize $S = \sum_{i=1}^{n} (y_i - \theta)^2$ we need to compute its derivative with respect to θ and equate to zero:

$$\frac{\partial S}{\partial \theta} = -2 \sum_{i=1}^{n} (y_i - \theta) = -2(n\bar{y} - n\theta)$$

and, putting a hat on θ,

$$-2(n\bar{y} - n\hat{\theta}) = 0 \quad \Longrightarrow \quad \hat{\theta} = \bar{y}.$$

It remains to verify that $\theta = \hat{\theta} = \bar{y}$ is indeed a minimum of S. We use the second derivative test:

$$\frac{\partial^2 S}{\partial \theta^2} = 2n > 0$$

Thus S indeed attains a minimum at $\theta = \hat{\theta} = \bar{y}$, and so $\hat{\theta} = \bar{y}$ is the least squares estimate (l.s.e) of θ.

Exercise 3 (Section 2.4)**.**

1. Let $Y_1 \sim N(1,3)$ and $Y_2 \sim N(2,5)$ be independent random variables. Find the distributions of $W_1 = Y_1 + 2Y_2$ and $W_2 = 4Y_1 - Y_2$.
2. Let $Y_1 \sim N(0,1)$ and $Y_2 \sim N(1,9)$ be independent random variables. Find the distribution of $W = Y_1^2 + (Y_2 - 1)^2/9$.
3. Let $Y_1 \sim N(0,1)$ and $Y_2 \sim N(3,4)$ be independent random variables. Set $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ (Y_2-3)/2 \end{pmatrix}$. Find the distribution of $\boldsymbol{Y}^T \boldsymbol{Y}$.
4. Let $Y_1, \ldots, Y_n$ be independent random variables each with the distribution $N(\mu, \sigma^2)$. Find the distribution of $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$.

Solution.

1. We need to use the property that if $Y_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$ are independent random variables, then

$$Z = \sum_{i=1}^n a_i Y_i \sim N\Big(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\Big)$$

   where $a_1, \ldots, a_n$ are arbitrary constants.

   We have $Y_1 \sim N(1,3)$ i.e. $\mu_1 = 1$, $\sigma_1^2 = 3$ and $Y_2 \sim N(2,5)$ i.e. $\mu_2 = 2$, $\sigma_2^2 = 5$. Thus

$$W_1 = Y_1 + 2Y_2 = 1Y_1 + 2Y_2 \sim N\big(\underbrace{1\cdot 1 + 2\cdot 2}_{1\cdot\mu_1 + 2\cdot\mu_2}, \underbrace{1^2\cdot 3 + 2^2\cdot 5}_{1^2\cdot\sigma_1^2 + 2^2\cdot\sigma_2^2}\big) = N\big(5, 23\big),$$

$$W_2 = 4Y_1 - Y_2 = 4Y_1 + (-1)Y_2 \sim N\big(\underbrace{4\cdot 1 - 1\cdot 2}_{4\cdot\mu_1 - 1\cdot\mu_2}, \underbrace{4^2\cdot 3 + (-1)^2\cdot 5}_{4^2\cdot\sigma_1^2 + (-1)^2\cdot\sigma_2^2}\big) = N\big(2, 53\big).$$

2. Recall that if $Z \sim N(\mu, \sigma^2)$ then $(Z-\mu)/\sigma \sim N(0,1)$. Thus $(Y_2 - 1)/3 \sim N(0,1)$, since $Y_2 \sim N(1,9)$. Then using definition of the Chi-squared distribution (see p.13 of lecture notes) we find that

$$W = \underbrace{Y_1^2}_{\sim N(0,1)^2} + \underbrace{\big((Y_2-1)/3\big)^2}_{\sim N(0,1)^2} \sim \chi_2^2.$$

3. Again, we notice that $(Y_2 - 3)/2 \sim N(0,1)$. Thus

$$\boldsymbol{Y}^T\boldsymbol{Y} = \underbrace{Y_1^2}_{\sim N(0,1)^2} + \underbrace{\big((Y_2-3)/2\big)^2}_{\sim N(0,1)^2} \sim \chi_2^2.$$

4. This question is similar to part 1:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N\Big(\sum_{i=1}^{n}\frac{\mu}{n}, \sum_{i=1}^{n}\frac{\sigma^2}{n^2}\Big) = N\Big(\mu, \frac{\sigma^2}{n}\Big).$$

Exercise 4 was not discussed in class.

Exercise 4 (Sections 2.4, 2.5). Suppose that $\boldsymbol{y} \sim N_4(\boldsymbol{\mu}, V)$, a random 4×1 vector, i.e. $\boldsymbol{y} = (y_1, y_2, y_3, y_4)^T$, described by a multivariate normal distribution with

$$\boldsymbol{\mu} = \begin{pmatrix} -4 \\ 2 \\ 5 \\ -1 \end{pmatrix}, \qquad V = \begin{pmatrix} 8 & 0 & -1 & 0 \\ 0 & 3 & 0 & 2 \\ -1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 7 \end{pmatrix}.$$

1. Find the trace of the matrix V.
2. Find the distribution of y_2 and of y_3.
3. Find the distribution of $z = y_1 + 2y_2 - y_4$.
4. Find the joint distribution of y_1 and y_3, i.e. the multivariate normal dist. of $\begin{pmatrix} y_1 \\ y_3 \end{pmatrix}$.
5. Find the joint distribution of y_1, y_2 and $\frac{1}{2}(y_2 + y_4)$.
6. Which pairs of random variables are independent, i.e. are y_1 and y_2 independent? are y_1 and y_3 independent? and so on.

Solution. We will need to use the following property. Let $\boldsymbol{y} = (y_1, \dots, y_r)^T \sim N_r(\mu, V)$, a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_r)^T$ and $r \times r$ variance-covariance matrix V. Let $\boldsymbol{a} = (a_1, \dots, a_p)^T$ be a constant $p \times 1$ vector, and let B be a constant $p \times r$ matrix. Then

$$\boldsymbol{a} + B\boldsymbol{y} \sim N_p\big(\boldsymbol{a} + B\boldsymbol{\mu}, BVB^T\big).$$

1. $\operatorname{tr} V = 8 + 3 + 5 + 7 = 23$
2. Notice that the mean and variance of y_i is μ_i and V_{ii}. Thus

$$y_2 \sim N(\mu_2, V_{22}) = N(2,3), \qquad y_3 \sim N(\mu_3, V_{33}) = N(5,5).$$

3. Choose $B = (1,2,0,-1)$. Then

$$B\boldsymbol{y} = (1,2,0,-1)\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = y_1 + 2y_2 - y_4 = z$$

and

$$B\boldsymbol{\mu} = (1,2,0,-1)\begin{pmatrix} -4 \\ 2 \\ 5 \\ -1 \end{pmatrix} = -4+4+1 = 1$$

and

$$BVB^T = (1,2,0,-1)\begin{pmatrix} 8 & 0 & -1 & 0 \\ 0 & 3 & 0 & 2 \\ -1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 7 \end{pmatrix}(1,2,0,-1)^T = 19.$$

Hence

$$z = B\boldsymbol{y} = N(B\boldsymbol{\mu}, BVB^T) = N(1,19).$$

4. Choose $B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$ so that

$$B\boldsymbol{y} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_3 \end{pmatrix}$$

and

$$B\boldsymbol{\mu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} -4 \\ 2 \\ 5 \\ -1 \end{pmatrix} = \begin{pmatrix} -4 \\ 5 \end{pmatrix}$$

and

$$BVB^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 8 & 0 & -1 & 0 \\ 0 & 3 & 0 & 2 \\ -1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 7 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 8 & -1 \\ -1 & 5 \end{pmatrix}$$

Hence

$$z = B\boldsymbol{y} = N_2\left(\begin{pmatrix} -4 \\ 5 \end{pmatrix}, \begin{pmatrix} 8 & -1 \\ -1 & 5 \end{pmatrix}\right).$$

5. Choose $B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}$ so that

$$B\boldsymbol{y} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ 1/2(y_2 + y_4) \end{pmatrix}$$

and

$$B\boldsymbol{\mu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} -4 \\ 2 \\ 5 \\ -1 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \\ 1/2 \end{pmatrix}$$

and

$$BVB^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 8 & 0 & -1 & 0 \\ 0 & 3 & 0 & 2 \\ -1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 3 & 5/2 \\ 0 & 5/2 & 7/2 \end{pmatrix}$$

Hence

$$z = B\boldsymbol{y} = N_2\left(\begin{pmatrix} -4 \\ 2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} 8 & 0 & 0 \\ 0 & 3 & 5/2 \\ 0 & 5/2 & 7/2 \end{pmatrix} \right).$$

6. The pairs (y_1, y_2), (y_2, y_3), (y_3, y_4) and (y_1, y_4) are pairs of independent random variables, because their respective covariances $V_{12} = V_{21} = V_{23} = V_{32} = V_{34} = V_{43} = V_{14} = V_{41} = 0$.