

## Solutions Chapter 1 (Descriptive Statistics) – Mathematical Statistics

1.

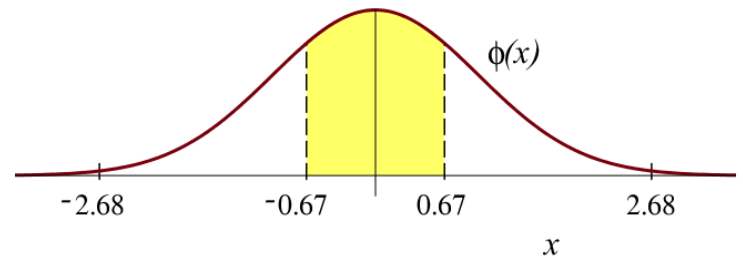
	Sample mean $\bar{x}$	Sample median $M$	Sample standard deviation $s$	Sample variance $s^2$
a.	2.5	3	3.15	9.9
b.	3.08	3	1.19	1.41
c.	49.6	49	8.77	76.93

2. The z-score of 40 is  $-2$ , so  $\frac{40 - \bar{x}}{s} = -2$ . And the z-score of 90 is 3, so  $\frac{90 - \bar{x}}{s} = 3$ .  
This information leads to two equations with the unknown  $\mu$  and  $\sigma$ :  $40 - \bar{x} = -2s$  and  $90 - \bar{x} = 3s$ .

Solving the equations we find:  $s = 10$  and  $\bar{x} = 60$ .

(But you can also argue that the difference between the observations 90 and 40 is 5 standard deviations (2 to the left and 3 to the right of  $\bar{x}$ ), so  $s = \frac{90-40}{5} = 10$  and the sample mean is  $40 + 2 \times 10 = 60$ )

The quartiles and the outliers



3.

- a.  $P(Z \leq Q_3) = 0.75$  if  $Q_3 \approx 0.67$  (N(0,1)-table)

Because of symmetry we have:

$$Q_1 = -Q_3 \approx -0.67$$

- b.  $IQR = Q_3 - Q_1 = 1.34$ , so

$$\begin{aligned} (Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR) \\ = (-0.67 - 1.5 \times 1.34, 0.67 + 2.01) \\ = (-2.68, +2.68) \end{aligned}$$

- c. Using symmetry:  $2 \times P(Z \geq Q_3 + 1.5 \times IQR) = 2 \times P(Z \geq 2.68) = 2[1 - 0.9963] = 0.74\%$

- d. The cdf of the exponential distribution is  $F(x) = 1 - e^{-\lambda x}$ ,  $x \geq 0$  ( $= \int_{-\infty}^x f(u) du$ )

$$F(x) = 0.75 \text{ if } x = -\frac{\ln(0.25)}{\lambda} = \frac{\ln(4)}{\lambda}. \text{ And } F(x) = 0.25 \text{ for } x = -\frac{\ln(0.75)}{\lambda} = \frac{\ln(\frac{4}{3})}{\lambda},$$

$$\text{So } IQR = Q_3 - Q_1 = \frac{\ln(4)}{\lambda} - \frac{\ln(\frac{4}{3})}{\lambda} = \frac{\ln(3)}{\lambda}$$

$$\begin{aligned} (Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR) &= \left( \frac{\ln(\frac{4}{3})}{\lambda} - 1.5 \cdot \frac{\ln(3)}{\lambda}, \frac{\ln(4)}{\lambda} + 1.5 \cdot \frac{\ln(3)}{\lambda} \right) \\ &= \left( \frac{\ln(\frac{4}{3} \cdot \frac{1}{3^{1.5}})}{\lambda}, \frac{\ln(4 \cdot 3^{1.5})}{\lambda} \right) = \left( \frac{\ln(\frac{4}{27} \sqrt{30})}{\lambda}, \frac{\ln(12\sqrt{3})}{\lambda} \right) \end{aligned}$$

The left bound is negative so the probability of an outlier is

$$P\left(X \geq \frac{\ln(12\sqrt{3})}{\lambda}\right) = e^{-\lambda \cdot \frac{\ln(12\sqrt{3})}{\lambda}} = \frac{1}{12\sqrt{3}} \approx 4.8\% : \text{ larger than for a normal distribution (0.74\%).}$$

4.

- a. First we will determine the 25<sup>th</sup> and the 75<sup>th</sup> percentile:  $Q_1 = x_{(10)} = 141$  and  $Q_3 = x_{(29)} = 154$ ,  
since 25% (75%) of  $n = 38$  is 9.5 (28.5).

The Inter Quartile Range  $IQR = 154 - 141 = 13$

Possible outliers are outside  $(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR) = (121.5, 173.5)$ :

we have one (potential) **outlier: 121**

- b. 1. The numerical summary: sample skewness coefficient = -0.41 deviates somewhat from 0, the reference value of the normal distribution. The kurtosis = 3.09 is close to the reference value 3.  
2. The histogram reveals a distribution which follows the normal distribution roughly, though there

seem to be two peaks. But the histogram is roughly symmetric and mound shaped.

3. The **normal** Q-Q plot shows points around the straight line  $y = x$ : normality seems reasonable. The total “verdict” based on these three aspects, is that the normal distribution can be assumed a model of these observations.

5.

a.  $E(X_i) = \frac{1}{2}$  (symmetry),  $E(X_i^2) = \int_0^1 x^2 dx = \left[ \frac{1}{3} x^3 \right]_{x=0}^{x=1} = \frac{1}{3}$ , so  $var(X) = E(X^2) - (EX)^2 = \frac{1}{12}$ .  
 $\gamma_2 = \frac{E(X-\mu)^4}{\sigma^4}$ , where  $\sigma^4 = (var(X))^2 = \frac{1}{144}$  and

$$E(X - \mu)^4 = \int_0^1 \left(x - \frac{1}{2}\right)^4 \cdot 1 dx = \left[ \frac{1}{5} \left(x - \frac{1}{2}\right)^5 \right]_{x=0}^{x=1} = \frac{1}{5} \left[ \frac{1}{2^5} - \left(-\frac{1}{2^5}\right) \right] = \frac{1}{80}$$

Hence  $\gamma_2 = \frac{E(X-\mu)^4}{\sigma^4} = \frac{144}{80} = 1.8$ .

b.  $F_{X_1}(x) = \int_0^x 1 du = [u]_{u=0}^{u=x} = x$  (if  $0 \leq x \leq 1$ ).  $F_{X_1}(x) = 0$  for  $x < 0$  and  $F_{X_1}(x) = 1$  for  $x > 1$ .

c.  $F_Y(y) = P(Y \leq y) = P(-2 \ln(X_1) \leq y) = P(\ln(X_1) \geq -\frac{1}{2}y)$   
 $= P(X_1 \geq e^{-\frac{1}{2}y}) = 1 - F_{X_1}(e^{-\frac{1}{2}y}) = 1 - e^{-\frac{1}{2}y}$ , or  $0 \leq e^{-\frac{1}{2}y} \leq 1$ , implying that  $y \geq 0$ .

Conclusion:  $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{2} e^{-\frac{1}{2}y}$  ( $y \geq 0$ ).

This is the exponential density with parameter  $\lambda = \frac{1}{2}$ , so  $E(Y) = \frac{1}{\lambda} = 2$ .

d.  $P(Z \leq z) = P(\max(X_1, \dots, X_n) \leq z) = P(X_1 \leq z \text{ and } X_2 \leq z \text{ and } \dots \text{ and } X_n \leq z)$   
 $\stackrel{\text{ind.}}{=} P(X_1 \leq z) \cdot \dots \cdot P(X_n \leq z) = z^n$ , if  $0 \leq z \leq 1$

So  $f_Z(z) = \frac{d}{dz} [z^n] = n z^{n-1}$ , for  $0 \leq z \leq 1$  (and  $f_Z(z) = 0$  elsewhere)

$$E(Z) = \int_{-\infty}^{\infty} z f(z) dz = \int_0^1 z \cdot n z^{n-1} dz = \left[ \frac{n}{n+1} z^{n+1} \right]_{z=0}^{z=1} = \frac{n}{n+1}$$

$$E(Z^2) = \int_0^1 z^2 \cdot n z^{n-1} dz = \left[ \frac{n}{n+2} z^{n+2} \right]_{z=0}^{z=1} = \frac{n}{n+2}$$
, so  $ar(Z) = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{(n+2)(n+1)^2}$ .

6. (We are using the notation  $X$  instead of  $X_1$  in a.)

a.  $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \dots \text{integr. by parts} \dots = [x \cdot -e^{-\lambda x}]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx$   
 $= 0 + \frac{1}{\lambda}$

$$E(X^2) = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx = \dots = \frac{2}{\lambda^2}$$
, so  $var(X) = E(X^2) - (EX)^2 = \frac{1}{\lambda^2}$  and  $\sigma = \frac{1}{\lambda}$ .

$$E(X^3) = \int_0^{\infty} x^3 \cdot \lambda e^{-\lambda x} dx = [x^3 \cdot -e^{-\lambda x}]_0^{\infty} + \int_0^{\infty} 3x^2 e^{-\lambda x} dx = 0 + \frac{3}{\lambda} \cdot E(X^2) = \frac{6}{\lambda^3}$$
.

$$E(X - \mu)^3 = E(X^3 - 3\mu \cdot X^2 + 3\mu^2 \cdot X - \mu^3) = E(X^3) - 3\mu \cdot E(X^2) + 2\mu^3 = \frac{6 - 6 + 2}{\lambda^3} = \frac{2}{\lambda^3}$$
.

So  $\gamma_1 = \frac{E(X-\mu)^3}{\sigma^3} = \frac{2/\lambda^3}{(\frac{1}{\lambda})^3} = 2$ .

b. For  $n = 100$  and  $\lambda = 2$ ,  $S$  is approximately (CLT)  $N(n\mu, n\sigma^2) = N(50, 25)$ .

So  $P(S > 55) = 1 - P(S \leq 55) \stackrel{\text{CLT}}{\approx} 1 - \Phi\left(\frac{55-50}{\sqrt{25}}\right) = 1 - \Phi(1) = 1 - 0.8413 = 15.87\%$ .

c. For  $n = 50$  and  $\lambda = 2$ ,  $\bar{X}$  is approximately  $N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(25, \frac{1}{200}\right)$ .

$$P(\bar{X} > 0.55) \stackrel{\text{CLT}}{\approx} 1 - \Phi\left(\frac{0.55-0.50}{\sqrt{\frac{1}{200}}}\right) = 1 - \Phi(0.71) = 0.2389$$
.

d.  $F_M(m) = P(\min(X_1, \dots, X_n) \leq m) = 1 - P(\min(X_1, \dots, X_n) \geq m)$   
 $= 1 - P(X_1 \geq m \text{ and } \dots \text{ and } X_n \geq m)$

$\stackrel{\text{ind.}}{=} 1 - P(X_1 \geq m) \cdot \dots \cdot P(X_n \geq m) = 1 - [e^{-\lambda m}]^n = 1 - e^{-n\lambda m} \quad (m \geq 0)$   
 $f_M(m) = n\lambda e^{-n\lambda m} \quad (\geq 0)$ , so  $M$  is exponentially distributed with parameter  $n\lambda$ .  
 For  $n = 10$  and  $\lambda = 2$ ,  $E(M) = \frac{1}{10 \cdot 2} = 0.05$ .

**7. a.**  $X + Y \sim \text{Poisson}(\mu = 3 + 4)$ .

You can state this if you know this property: if  $X$  and  $Y$  are independent and both Poisson-distributed with parameters  $\mu_X$  and  $\mu_Y$ , resp., then  $Z = X + Y$  has a Poisson distribution with parameter  $\mu_X + \mu_Y$ . This is proven as follows (using the **convolution sum**, here with  $\mu_X = 3$  and  $\mu_Y = 4$ ):

$$P(Z = n) \stackrel{\text{ind.}}{=} \sum_{k=0}^n P(X = k) \cdot P(Y = n - k) \quad (\text{using the assumed independence of } X \text{ and } Y).$$

$$= \sum_{k=0}^n \frac{3^k e^{-3}}{k!} \cdot \frac{4^{n-k} e^{-4}}{(n-k)!} = e^{-7} \sum_{k=0}^n \frac{1}{k! (n-k)!} 3^k 4^{n-k} = \frac{e^{-7}}{n!} \sum_{k=0}^n \binom{n}{k} 3^k 4^{n-k}$$

In the last summation we recognize Newton's Binomial Theorem:  $\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a + b)^n$ :

$$P(Z = n) = \frac{e^{-7}}{n!} \cdot (3 + 4)^n = \frac{7^n e^{-7}}{n!}, \quad \text{for } n = 0, 1, 2, \dots$$

Alternatively one can use moment generating functions: for the Poisson distribution of  $N$  we have:

$$M_N(t) = E(e^{tN}) = \sum_{n=0}^{\infty} e^{tn} \cdot \frac{\mu^n}{n!} e^{-\mu} = \sum_{n=0}^{\infty} \frac{(e^t \mu)^n}{n!} e^{-\mu} = e^{-\mu} e^{e^t \mu} \sum_{n=0}^{\infty} \frac{(e^t \mu)^n}{n!} e^{-e^t \mu} = e^{-\mu(1-e^t)}$$

For independent  $X \sim \text{Poisson}(\mu_1)$  and  $Y \sim \text{Poisson}(\mu_2)$ , we have

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{-\mu_1(1-e^t)} \cdot e^{-\mu_2(1-e^t)} = e^{-(\mu_1+\mu_2)(1-e^t)}$$

Conclusion:  $X + Y \sim \text{Poisson}(\mu_1 + \mu_2)$  (proof is easily extended to  $n$  independent  $X_i \sim \text{Poisson}(\mu_i)$ )

**b.** Since  $X$  and  $Y$  are independent we can apply the **convolution integral**, but we have to take into account that  $f_X(x)$  and  $f_Y(y)$  are both 1 if both  $x$  and  $y$  are between 0 and 1, otherwise  $f_X(x) \cdot f_Y(y) = 0$ .

(Sketch the integration area in the  $XY$ -plane, see where the line  $x + y = z$  intersects the unit square)

Distinguishing  $z = x + y$  between 0 and 1, between 1 and 2 and all other values we find:

$$0 \leq z \leq 1: f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_0^z 1 \cdot 1 dx = [x]_{x=0}^z = z$$

$$1 < z \leq 2: f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_{z-1}^1 1 \cdot 1 dx = [x]_{x=z-1}^1 = 1 - (z-1) = 2 - z$$

$$\text{So: } f(x) = \begin{cases} z & \text{if } 0 \leq z \leq 1 \\ 2 - z & \text{if } 1 \leq z \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

**c.**  $X + Y \sim B(m + n, p)$ .

This is a property of the binomial distribution which can be proven as follows:

Suppose that  $X_1, \dots, X_m, X_{m+1}, \dots, X_{m+n}$  are independent alternatives with success probability  $p$ , then

$$X = \sum_{i=1}^m X_i \sim B(m, p), \quad Y = \sum_{i=m+1}^{m+n} X_i \sim B(n, p) \quad \text{and} \quad X + Y = \sum_{i=1}^{m+n} X_i \sim B(m + n, p).$$

**d.**  $X + Y \sim N(20 + 30, 81 + 144)$ .

According to a well known property of the normal distribution.

(Using the moment generating of a  $N(\mu, \sigma^2)$ -distribution,  $M(t) = e^{\mu t + \sigma^2 t^2 / 2}$ , it is easily shown that the sum  $X + Y$  of two independent  $X \sim N(\mu_1, \sigma_1^2)$  and  $Y \sim N(\mu_2, \sigma_2^2)$  has a  $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ -distribution)

**e.**  $X \sim \text{Exp}(\lambda = 2)$  and  $Y \sim \text{Exp}(\lambda = 3)$  are independent, so:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \stackrel{z > 0}{=} \int_0^z 2e^{-2x} \cdot 3e^{-3(z-x)} dx$$

$$= 6e^{-3z} \int_0^z e^x dx = 6e^{-3z} \cdot [e^x]_{x=0}^z = 6e^{-3z} \cdot (e^z - 1), \quad \text{for } z > 0$$

And  $f_{X+Y}(z) = 0$  if  $z < 0$ .

This is a density function since  $f_{X+Y}(z) \geq 0$  and

$$\int_{-\infty}^{\infty} f_{X+Y}(z) dz = \int_0^{\infty} (6e^{-2z} - 6e^{-3z}) dz = [-3e^{-2z} + 2e^{-3z}]_{x=0}^{\infty} = 0 - (-3 + 2) = 1.$$