Mathematical Statistics
Week 5 exercises

1. Let us have a realisation of a random sample $x_1, \dots, x_n$ from $Be(p)$, where the parameter p is an unknown parameter. Consider the following two estimates of p:

$$\hat{p}_1 = \overline{x} \quad \text{and} \quad \hat{p}_2 = \frac{x_1 + x_2}{2}.$$

   (a) Are the estimates $\hat{p}_1$ and $\hat{p}_2$ unbiased?
   (b) Which of the estimates is more efficient?
   (c) Are the estimates $\hat{p}_1$ and $\hat{p}_2$ consistent?

2. Two methods are available for determining a certain chemical quantity p. Using these methods, we have obtained two independent estimates $\hat{p}_1$ and $\hat{p}_2$, which are both unbiased. The standard deviations of the estimates are 0.4 and 0.6, respectively. Combine the estimates into a single unbiased estimate by forming the expression

$$\hat{p} = \alpha\hat{p}_1 + (1-\alpha)\hat{p}_2, \quad \text{where } 0 \le \alpha \le 1.$$

   (a) Prove that $\hat{p}$ is an unbiased estimate of p.
   (b) Find the value of α for which the variance of $\hat{p}$ is as small as possible.

3. Show that if we assume the random sample to have a distribution $N(\mu, \sigma^2)$, where μ is known and σ is unknown, then the estimate

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

   is an unbiased estimate for the parameter σ^2, and if the estimate

$$\hat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$$

   is used, we obtain a biased estimate for the parameter σ^2. Find the bias.

4. For two data sets $x_1, \dots, x_4$ and $y_1, \dots, y_5$ the arithmetic means and variances are

$$\overline{x} = 1.38, \quad \overline{y} = 2.10,$$

$$s_x^2 = 0.0552, \quad s_y^2 = 0.0792.$$

   Compute the arithmetic mean and variance, assuming that all nine values are regarded as one data set.

5. Let $X \sim Bin(m, p)$, where p is unknown. We have an observation x on X. Find an unbiased estimate of the variance $Var(X)$.

6. Let $x = 16$ be an observation on $X \sim Bin(25, p)$.
   (a) Estimate p.
   (b) Find the standard deviation of the estimate.

7. Non-parametric statistics - In this course we will mainly deal with **parametric statistics**, where somewhere along the way we assume that our data or statistic follows or is drawn from a given parametric family of probability distributions (that is determined by fixing parameters, and hence the name). For example, in applying t-test we assume that our data is normally distributed with certain mean and variance.

   If this assumption is not justified, then non-parametric statistics - a special branch of statistics - comes to help and loosens these assumptions. But be aware, results found using non-parametric statistics are not that powerful (i.e. we need a lot more data to reach the same level of confidence when using parametric statistics).

   A fun video that explains the difference once more.

$\boxed{1}$

realised random sample from

$x_1, \dots, x_n \sim Be(p)$

$\hat{P}_1 = \overline{X}$ $\qquad \hat{P}_2 = \frac{x_1 + x_2}{2}$

according to Bernoulli distribution

p $EX = 1 \cdot p + 0(1-p)$

$= p$

a)

$$E[\hat{P}_1] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum^{n} E[X_i] = p$$

$$E[\hat{P}_2] = E\left[\frac{X_1 + X_2}{2}\right] = \frac{2p}{2} = p$$

$Var X = E[X^2] - (EX)^2$

$= p - p^2$

b)

$$Var\,\hat{P}_1 = \frac{1}{n^2} Var \sum^{n} X_i = \frac{1}{n^2}\sum^{n} Var X_i = \frac{\cancel{n}\,p(1-p)}{n^{\cancel{2}}}$$

$$Var\,\hat{P}_2 = \frac{1}{4}\big(p(1-p) + p(1-p)\big) = \frac{1}{2}p(1-p)$$

$\hat{P}_1$ is more efficient

c)

$\hat{P}_2$ is not consisdent by (2)

$\boxed{2}$

a)

$$E[\hat{P}] = E[\alpha \hat{P}_1 + (1-\alpha)\hat{P}_2] = \alpha E\hat{P}_1 + (1-\alpha) E\hat{P}_2 =$$

$$= \alpha p + (1-\alpha)p = p$$

b) $$Var[\hat{P}] = Var[\alpha \hat{P}_1 + (1-\alpha)\hat{P}_2] =$$

$$= \alpha^2 Var \hat{P}_1 + (1-\alpha)^2 Var \hat{P}_2 = \alpha^2 0{,}4^2 + (1-\alpha)^2 0{,}6^2$$

next derivative

$$2 \cdot 0{,}4^2 \alpha - 2 \cdot 0{,}6^2 + 2 \cdot 0{,}6^2 \alpha = 0$$

$$0{,}32\alpha - 0{,}72 + 0{,}72\alpha = 0$$

$$\alpha = \frac{0{,}72}{1{,}04} \approx 0{,}6923$$

solution

$\alpha = 0{,}6923$

3 Random sample has distribution $N(\mu, \sigma^2)$ where μ is known and σ^2 is not.

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

show $\hat{\sigma}_1^2$ is unbiased

$$E[\hat{\sigma}_1^2] = \frac{1}{n} E \sum (x_i - \mu)^2 = \frac{1}{n} \sum E(x_i - \mu)^2 =$$

$$= \frac{n}{n} \sigma^2$$

?

Do: need to expand!?

$\sigma^2 = EX^2 - (EX)^2$

$$= \frac{1}{n} \left(\sum EX_i^2 - 2\mu \sum EX_i + n\mu^2 \right) =$$

$$= \frac{1}{n} \left(n(\sigma^2 + \mu^2) - 2n\mu^2 + n\mu^2 \right) =$$

$$= \frac{1}{n} \left(n\sigma^2 + n\mu^2 - 2n\mu^2 + n\mu^2 \right) = \sigma^2$$

show $\hat{\sigma}_2^2$ is biased

$$E[\hat{\sigma}_2] = \frac{1}{n-1} E \sum (x_i - \mu)^2 = \frac{n}{n-1} \sigma^2$$

$$= \frac{n}{n-1} E[\hat{\sigma}_1] = \frac{n}{n-1} \sigma^2$$

4

$x = (x_1, \dots x_4)$

$$\bar{x} = \frac{x_1 + \dots + x_4}{4} = 1,38$$

$$\bar{y} = \frac{y_1 + \dots + y_5}{5} = 2,10$$

$$S_x^2 = \frac{1}{3} \sum_{i=1}^{4} (x_i - \bar{x})^2 = \frac{1}{3}\left(EX^2 - (E\bar{X})^2\right)$$

$$S_y^2 = \frac{1}{4} \sum_{i=1}^{5} (y_i - \bar{y})^2 = \frac{1}{4}\left(EY^2 - (EY)^2\right)$$

Assume all 9 items belong to a single dataset z.

$$\bar{z} = \frac{4\bar{x} + 5\bar{y}}{9} = \frac{16,02}{9} = 1,78$$

$$S_z^2 = \frac{1}{8}\left(Ez^2 - (\bar{z})^2\right) = \frac{1}{8}\left(\sum^{4} (x_i - \bar{z})^2 + \sum^{5} (y_i - \bar{z})^2\right)$$

$$= \frac{1}{8}\left(\sum (x_i - \bar{x} + \bar{x} - \bar{z})^2 + \sum (y_i - \bar{y} + \bar{y} - \bar{z})^2\right) =$$

same for y

$$= \frac{1}{8}\left(\sum^{4} (x_i - \bar{x})^2 + \cancel{2(\bar{x} - \bar{z}) \sum^{4} (x_i - \bar{x})}^{0} + 4(\bar{x} - \bar{z})^2 + \text{—''—}\right.$$

showed below

$$= \frac{1}{8}\left(3S_x^2 + 4(\bar{x} - \bar{z})^2 + 4S_y^2 + 5(\bar{y} - \bar{z})^2\right)$$

$$= 2(\bar{x}-\bar{z})\sum^{4}(x_i-\bar{x}) = 2(\bar{x}-\bar{z})(x_1-\bar{x}+x_2-\bar{x}+\dots)$$

$$= 2(\bar{x}-\bar{z})\left(\sum x_i - 4\bar{x}\right) = 2(\bar{x}-\bar{z})\left(\sum x_i - \frac{4\sum^{4} x_i}{4}\right) =$$

$$= 0$$

$$= \frac{1}{8}\left(3S_x^2 + 4(\bar{x}-\bar{z})^2 + 4S_y^2 + 5(\bar{y}-\bar{z})^2\right) =$$

$$= \frac{1}{8}\left(3\cdot 0{,}0552 + 4(1{,}38-1{,}78)^2 + 4\cdot 0{,}0792 + 5(2{,}1-1{,}78)^2\right) =$$

$$= \frac{1}{8}\left(0{,}1656 + 0{,}64 + 0{,}3168 + 0{,}512\right) =$$

$$= \frac{1}{8}\cdot 1{,}6344 = 0{,}2043$$

$\boxed{5}$

$$X \sim Bin(m, p) = \binom{m}{x} p^x (1-p)^{m-x}$$

$$EX = mp \qquad VX = mpq = mp(1-p)$$

Find an unbiased estimate for $Var\,X$

We have to estimate p with $\hat{p}$

we can estimate $\hat{p} = \frac{x}{m}$ e.g. number of heads out of m trials

$$E[\hat{Var}\,X] = E\left[m\hat{p}(1-\hat{p})\right] =$$

$$= m(E\hat{p} - E\hat{p}^2) = m\left(E\frac{X}{m} - \left(Var\,\hat{p} + (E\hat{p})^2\right)\right) =$$

$$= m\left(p - \left(\frac{1}{m^2}Var X + \frac{1}{m^2}(EX)^2\right)\right) =$$

$$= m\left(p - \frac{1}{m^2}\left(mp(1-p) + (mp)^2\right)\right) =$$

$$= mp - p(1-p) - mp^2 = mp(1-p) - p(1-p)$$

$$= p(1-p)(m-1)$$

$$\hat{Var}(X)_{corr} = \frac{n}{n-1} \hat{Var}(X) = n p (1-p)$$

$\boxed{6}$

a)

$$\hat{p} = \frac{x}{n} = \frac{16}{25}$$

b)

$$\hat{Var} X = \frac{n}{n-1} \hat{Var}(X) = 25 \cdot \frac{16}{25} \cdot \left(1 - \frac{16}{25}\right) =$$

$$= 16 \cdot \frac{9}{25} = 5{,}76$$

$$\sqrt{\hat{Var} X} = 2{,}4$$