1. (a) Med hjälp av den givna sannolikhetsfunktionen och räkneregeln för väntevärdet av en funktion av en stokastisk variabel fås

$$E(\frac{1}{X}) = \sum_k \frac{1}{k} p_X(k) = \frac{1}{1} \cdot 0.5 + \frac{1}{2} \cdot 0.2 + \frac{1}{3} \cdot 0.3 = 0.7$$

(b) För $X \in R(0,1)$ är $E(X) = 1/2$ och $V(X) = 1/12$ (formelsamlingen). För att beräkna den sökta kovariansen behövs även $E(X^2)$ och $E(X^3)$. Den förra kan t.ex. fås ur $V(X) = E(X^2) - E(X)^2$, men här kan vi lika gärna räka ut $E(X^k)$ för $k = 0, 1, 2, \ldots$ (eftersom det är lika enkelt som $E(X^3)$)

$$E(X^k) = \int_{-\infty}^{\infty} x^k \cdot f_X(x)dx = \int_0^1 x^k \cdot 1\, dx = \left[\frac{x^{k+1}}{k+1}\right]_0^1 = \frac{1}{k+1}. \quad k = 0, 1, 2, \ldots$$

$$C(X, X^2) = E(X \cdot X^2) - E(X)E(X^2) = \frac{1}{4} - \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}$$

(c) Om $X \in Po(2)$ så är, enligt formelsamlingen, $E(X) = 2$. Sannolikheten blir

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F_X(2) = [\text{Tabell 5}] = 1 - 0.67668 \approx 0.32$$

Om däremot X är normalfördelad så är ju dess (kontinuerliga) fördelning symmetrisk kring väntevärdet, så sannolikheten att detta överskrids måste därmed vara 1/2.

2. (a) Då $X_i \in N(\mu, \sigma)$ kan vi skatta den sökta sannolikheten om vi räknar ut den med skattade värden på μ och σ.

$$\mu^* = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{9} \cdot 407.4 = 45.2667$$

$$\sigma^* = s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = [\text{miniräknare}] = 1.0689$$

$$P(X_i \geq 47) = 1 - P(X_i < 47) \approx 1 - \Phi(\frac{47 - 45.2667}{1.0689}) \approx 1 - \Phi(1.62) \approx 0.053$$

(b) Ett 95% konfidensintervall för μ blir, eftersom $\mu^* = \bar{X} \in N(\mu, \sigma/\sqrt{n})$),

$$I_\mu = \mu^* \pm t_{\alpha/2}(f)d(\mu^*) = \bar{x} \pm t_{0.025}(n-1)\frac{s}{\sqrt{n}} = 45.2667 \pm 2.31 \cdot \frac{1.0689}{\sqrt{9}} = [44.4, 46.1]$$

Anm. De uppmätta kondensatorerna var märkta ”473” vilket är en kodbeteckning för 47 nF. En kalibrering av kapacitansmätaren skulle kunna avslöja om vi skall skylla på mätaren eller kondensatorerna.

3. Enligt uppgiften skall en siffra, 11 i det här fallet, förekomma i en lottorad med sannolikheten $p_0 = \frac{1}{5}$. Harry misstänker att sannolikheten är högre än så och vill därför testa

$$H_0: \ p = p_0$$
$$H_1: \ p > p_0$$

Om vi låter $X =$ ”Antal rader där siffran 11 förekommer vid $n = 10$ dragningar” så är $X \in Bin(n, p)$ där p är sannolikheten att 11 förekommer i en rad. Vi har en observation $x = 6$ av X. Vi kan inte använda normalapproximation av X (eller $p^* = X/n$) för att utföra testet (eftersom $np^*(1-p^*) \not> 10$) men direktmetoden går bra. P-värdet blir

$$P = P(\text{Få det vi fått eller värre om } H_0 \text{ är sann}) = P(X \geq 6 \text{ om } X \in Bin(n, p_0)) = 1 - P(X \leq 5) =$$
$$= 1 - F_X(5) = [\text{Tabell 6}, n = 10, p = 0.20, x = 5] = 1 - 0.99363 = 0.0064$$

Eftersom P-värdet är mindre än t.ex. standardnivån $\alpha = 0.01$ kan H_0 förkastas på nivån 0.01, dvs baserat på de tio raderna är Harrys misstanke befogad.

Anm. Om Harry inte bara tittat på de tio senaste raderna utan t.ex. på det årets samtliga rader på Lotto 1 och 2 så långt (86 dragningar) så förekom siffran 11 som ordinarie vinstnummer vid 21 tillfällen (24%) vilket inte ger signifikant resultat (på nivån 0.05). Turnumren baserat på de observationerna är i stället 12 eller 34.

4. Vi har följande summor och kvadratsummor

$$n = 9, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{9}\cdot 72 = 8, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{9}\cdot 94.5 = 10.5$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 636 - \frac{1}{9}\cdot 72^2 = 60$$

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 1\,070 - \frac{1}{9}\cdot 94.5^2 = 77.75$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) =$$

$$= 823.7 - \frac{1}{9}\cdot 72 \cdot 94.5 = 67.7$$

(a) Regressionsparametrarna skattas med

$$\beta^* = \frac{S_{xy}}{S_{xx}} = \frac{67.7}{60} = 1.1283$$

$$\alpha^* = \bar{y} - \beta^*\bar{x} = 10.5 - 1.1283 \cdot 8 = 1.473$$

$$(\sigma^2)^* = s^2 = \frac{1}{n-2}Q_0 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{1}{7}\left(77.75 - \frac{67.7^2}{60}\right) = 0.1945$$

$$\sigma^* = \sqrt{0.0124} = 0.441$$

(b) Ett 95% konfidensintervall för β ges av

$$I_\beta = \beta^* \pm t_{p/2}(n-2)d(\beta^*) = \beta^* \pm t_{0.025}(7)\frac{\sigma^*}{\sqrt{S_{xx}}} = 1.1283 \pm 2.36 \cdot \frac{0.441}{\sqrt{60}} = [0.99, 1.26]$$

(c) x_0 kan lösas ut ur $y = \alpha^* + \beta^* x_0$ och blir

$$x_0 = \frac{y - \alpha^*}{\beta^*} = \frac{11.4 - 1.473}{1.128} = 8.80$$

(d) Om vi betraktar y_i som observationer av $Y_i \in N(\beta x_i, \sigma)$ fås MK-skattningen av β genom att minimera $Q(\beta)$ enligt

$$Q(\beta) = \sum_{i=1}^{n}(y_i - E(Y_i))^2 = \sum_{i=1}^{n}(y_i - \beta x_i)^2$$

$$\frac{dQ}{d\beta} = -2\sum_{i=1}^{n}(y_i - \beta x_i)x_i = -2\sum_{i=1}^{n} x_i y_i + 2\beta\sum_{i=1}^{n} x_i^2 = 0 \quad \Longrightarrow$$

$$\beta^*_{MK} = \frac{\sum_1^n x_i y_i}{\sum_1^n x_i^2} = \frac{823.7}{636} = 1.30$$

5. Täthetsfunktionen för en $Exp(1)$-fördelning är (enligt formelsamlingen) $f_{X_i}(x) = e^{-x},\ x \geq 0$ och därmed 0 för negativa x.

(a) Här skall vi bestämma $P(X_i + X_j < 2)$. Låt oss för enkelhets skull sätta $X = X_i$ och $Y = X_j$. Den sökta sannolikheten kan då beräknas som end dubbelintegral under linjen $x + y = 2$ i första kvadranten (rita gärna)

$$\begin{aligned}P(X+Y<2) &= \iint_{x+y<2} f_{X,Y}(x,y)\,dxdy = \int_0^2 e^{-x}\int_0^{2-x} e^{-y}\,dydx = \\ &= \int_0^2 e^{-x}\left[-e^{-y}\right]_0^{2-x} dx = \int_0^2 e^{-x}(1-e^{-(2-x)})dx = \int_0^2 (e^{-x}-e^{-2})dx = \\ &= \left[-e^{-x}-xe^{-2}\right]_0^2 = 1-e^{-2}-2e^{-2} \approx 0.594\end{aligned}$$

Alternativt kan man först räkna ut täthetsfunkionen för $Z = X + Y$ med hjälp av faltningsformeln och sedan integrera den fram till punkten 2. Observera att $f_X(x) = 0$ för negativa x och att $f_Y(z - x) = 0$ för då $x > z$ (rita dem)

$$\begin{aligned}f_Z(z) &= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx = \int_0^z e^{-x}\cdot e^{-(z-x)}\,dx = \int_0^z e^{-x-z+x}\,dx = \\ &= e^{-z}\,[x]_0^z = ze^{-z},\ z \geq 0 \\ P(Z<2) &= \int_{-\infty}^2 f_Z(z)\,dz = \int_0^2 ze^{-z}\,dz = \left[-ze^{-z}\right]_0^2 + \int_0^2 e^{-z}\,dz = -2e^{-2} + [-e^{-z}]_0^2 = \\ &= -2e^{-2} + 1 - e^{-2} = \approx 0.594\end{aligned}$$

(b) Låt Y vara antalet, av de 10, som är mindre än 1. Då är Y binomialfördelad, $Y \in Bin(n, p)$, där $n = 10$ och

$$p = P(X_i < 1) = \int_{-\infty}^1 f_{X_i}(x)\,dx = \int_0^1 e^{-x}\,dx = [-e^{-x}]_0^1 = 1 - e^{-1} \approx 0.6321$$

Den sökta sannolikheten blir

$$P(Y \leq 3) = \sum_{k=0}^3 p_Y(k) = \sum_{k=0}^3 \binom{10}{k} 0.6321^k (1-0.6321)^{10-k} \approx 0.0345$$

(c) Sannolikheten att en av dem är mindre än 3 är

$$P(X_i < 3) = [\text{samma som ovan}] = 1 - e^{-3} \approx 0.9502$$

För den största av fyra blir det sannolikheten att alla är mindre än 3

$$P(\max(X_1,\ldots,X_4) < 3) = P(X_1 \leq 3, X_2 \leq 3, X_3 \leq 3, X_4 \leq 3) = (1-e^{-3})^4 \approx 0.8152$$

(d) Här kan man återigen tänka sig en binomialfördelningsmodell eller en ffg-fördelning (eller geometrisk), eller betrakta den minsta av fem eller konstatera att den sökta sannolikheten är helt enkelt sannolikheten att fem av dem är mindre än ett som, med hjälp av (b), fås till

$$(1-e^{-1})^5 \approx 0.1009$$

6. (a) ML-skattningen, α^*, av α ges av det α som maximerar likelihoodfunktionen $L(\alpha)$.

$$\begin{aligned}L(\alpha) &= \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \sqrt{\frac{2}{\pi}}\cdot\frac{x_i^2}{\alpha^{3/2}}e^{-x_i^2/(2\alpha)} = \left(\frac{2}{\pi}\right)^{n/2}\cdot\alpha^{-3n/2}\cdot\prod_{i=1}^n x_i^2 e^{-x_i^2/(2\alpha)} = \\ \ln L(\alpha) &= \frac{n}{2}\ln\frac{2}{\pi} - \frac{3n}{2}\ln\alpha + \sum_{i=1}^n \ln x_i^2 - \frac{1}{2\alpha}\sum_{i=1}^n x_i^2 \\ \frac{d\ln L(\alpha)}{d\alpha} &= 0 - \frac{3n}{2\alpha} + 0 + \frac{1}{2\alpha^2}\sum_{i=1}^n x_i^2 = 0 \qquad \Longrightarrow \\ \alpha^* &= \frac{1}{3n}\sum_{i=1}^n x_i^2\end{aligned}$$

(b) För att avgöra om skattningen är väntevärdesriktig skall vi undersöka om $E(\alpha^*) = \alpha$. Eftersom skattningen är en linjärkombination av de kvadrerade observationerna, X_i^2, behöver vi ha väntevärdet av dem. Detta fås allmänt ur $\int x^2 f_{X_i}(x)\,dx$, men eftersom både väntevärde och varians för observationerna är givna i uppgiften behöver vi inte integrera utan får direkt ur $V(X_i) = E(X_i^2) - E(X_i)^2$:

$$E(X_i^2) = V(X_i) + E(X_i)^2 = \alpha(3 - 8/\pi) + (\sqrt{8\alpha/\pi})^2 = 3\alpha$$

$$E(\alpha^*) = E\left(\frac{1}{3n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{3n}\sum_{i=1}^{n} E(X_i^2) = \frac{1}{3n}\sum_{i=1}^{n} 3\alpha = \alpha$$

Skattningen är således väntevärdesriktig.