Examen 2015 — Corrigé

Probabilité et Statistiques SIC

Solution 1. (a) (2 points) A probability space for this experiment is $(\Omega, \mathcal{F}, \Pr)$, where the sample space of X is

$$\Omega = \{0, 1, 2, 3\},$$

the event space is the set of all subsets of Ω, that is

$$\begin{aligned}\mathcal{F} &= \{A : A = \cup_{i \in \mathcal{I}} \{X = i\},\ \mathcal{I} \subset \{0, 1, 2, 3\}\} \\ &= \{\emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{0, 1\}, \{0, 2\}, \{0, 3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \\ &\quad \{0, 1, 2\}, \{0, 1, 3\}, \{0, 2, 3\}, \{1, 2, 3\}, \{0, 1, 2, 3\}\},\end{aligned}$$

and the probability distribution $P : \mathcal{F} \to [0, 1]$ is such that, for all $A \in \mathcal{F}$, $\Pr(A) = \sum_{i \in \mathcal{I}} \Pr(X = i)$, with $\Pr(X = i) = C_3^i \times 1/8$ since $X \sim B(3, 0.5)$.

So,

$$\Pr(X = 2 \mid X \geq 2) = \frac{\Pr(\{X = 2\} \cap \{X \geq 2\})}{\Pr(X \geq 2)} = \frac{\Pr(X = 2)}{\Pr(X = 2) + \Pr(X = 3)} = \frac{\binom{3}{2} 1/8}{\binom{3}{2} 1/8 + \binom{3}{3} 1/8} = 3/4.$$

(b) (2 points) Let S denote the sum of the scores at a given roll. The event $\{X = x\}$ happens if the roller doesn't lose or win at the first $x - 1$ rolls and then wins or loses at the xth roll. Each roll is a Bernoulli trial with probability of 'success'

$$p = \Pr(S \in \{2, 3, 7, 11, 12\}) = \tfrac{1}{36} + \tfrac{2}{36} + \tfrac{6}{36} + \tfrac{2}{36} + \tfrac{1}{36} = \tfrac{1}{3}.$$

Therefore, the number of rolls X until the first win or loss follows a geometric distribution of parameter p, and mass function

$$\Pr(X = x) = p(1 - p)^{x-1}, \quad x = 1, 2, \ldots.$$

Therefore, we have

$$\Pr(\text{loss} \mid X = 3) = \frac{\Pr(\{\text{loss}\} \cap \{X = 3\})}{\Pr(\{X = 3\})} = \frac{\left(1 - \frac{1}{3}\right)^2 \times \left(\frac{1}{36} + \frac{2}{36} + \frac{1}{36}\right)}{\frac{1}{3} \times \left(1 - \frac{1}{3}\right)^2} = \tfrac{1}{3}.$$

Alternatively, noting that the probability to lose/win in a given roll is constant,

$$\Pr(\text{loss} \mid X = 3) = \Pr(S \in \{2, 3, 12\} \mid S \in \{2, 3, 7, 11, 12\}) = \frac{\frac{1}{36} + \frac{2}{36} + \frac{1}{36}}{\frac{1}{3}} = \tfrac{1}{3}.$$

(c) (2 points) The expected value of $X + Y$ is

$$\begin{aligned}\mathrm{E}(X + Y) &= \mathrm{E}(X) + \mathrm{E}(Y) \\ &= \mathrm{E}(X \mid Y = 0) \times \Pr(Y = 0) + \mathrm{E}(X \mid Y = 1) \times \Pr(Y = 1) + \mathrm{E}(Y) \\ &= \tfrac{7}{2} \times \tfrac{1}{2} + 0 + \tfrac{1}{2} = \tfrac{7}{4} + \tfrac{2}{4} = \tfrac{9}{4}.\end{aligned}$$

(d) **(2 points)** Let $Y = g(X)$ with $g(x) = 1/x$. Since g is monotonic with inverse $g^{-1}(y) = 1/y$ on (0, 1], then

$$f_Y(y) = f_X\{g^{-1}(y)\} \times \left|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right| = (1/y)^{-2} \times \frac{1}{y^2} = 1, \quad 0 < y \leq 1.$$

Alternatively,

$$F_X(x) = \int_1^x t^{-2}\,\mathrm{d}t = \left[-\tfrac{1}{t}\right]_1^x = 1 - x^{-1}, \quad x \geq 1.$$

So, for $0 < y \leq 1$ we have

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr\left(\tfrac{1}{X} \leq y\right) = \Pr\left(\tfrac{1}{y} \leq X\right) = 1 - \Pr\left(X < \tfrac{1}{y}\right) \\ &= 1 - F_X(1/y) = 1 - (1 - y) = y, \end{aligned}$$

which implies that $Y \sim U(0,1)$; Y has the standard uniform distribution.

(e) **(2 points)** We have $X_1 \sim \text{Pois}(\theta_1)$ and $X_2 \sim \text{Pois}(\theta_2)$. Then,

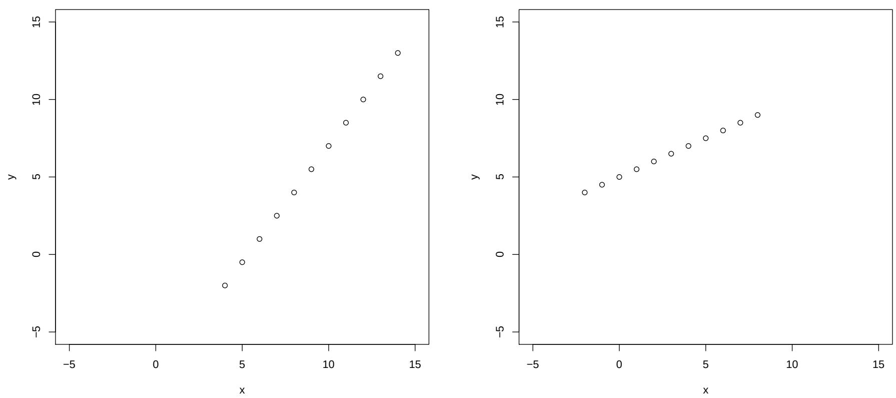
$$M_{X_1}(t) = \mathrm{E}(e^{tX_1}) = \sum_{x=0}^{\infty} e^{tx} e^{-\theta_1} \frac{\theta_1^x}{x!} = \sum_{x=0}^{\infty} e^{-\theta_1} \frac{(e^t\theta_1)^x}{x!} = e^{-\theta_1} \exp(\theta_1 e^t) = \exp\{\theta_1(e^t - 1)\},$$

where we have used the exponential series $\exp(z) = \sum_{n=0}^{\infty} z^n/n!$ for any $z \in \mathbb{R}$. Similarly, $M_{X_2}(t) = \exp\{\theta_2(e^t - 1)\}$, So,

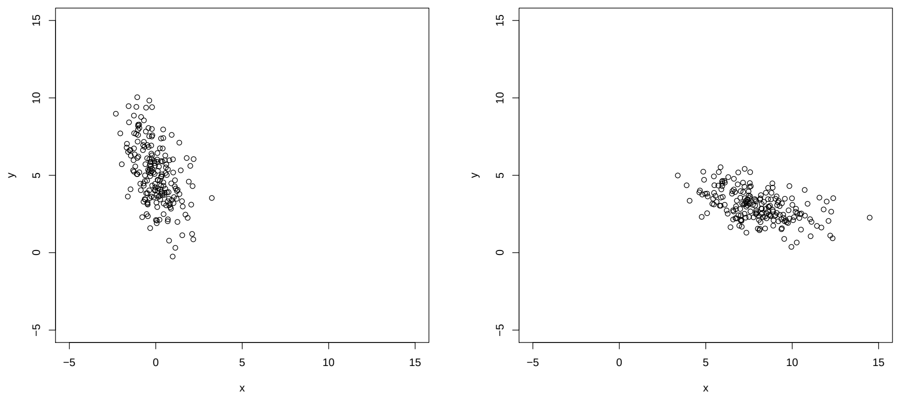
$$M_Z(t) = M_{X_1+X_2}(t) = M_{X_1}(t) \times M_{X_2}(t) = \exp\{(\theta_1 + \theta_2)(e^t - 1)\},$$

which is the moment-generating function of a Poisson distribution with parameter $\theta_1+\theta_2$. So Z follows a Poisson distribution with parameter $\theta_1 + \theta_2$.

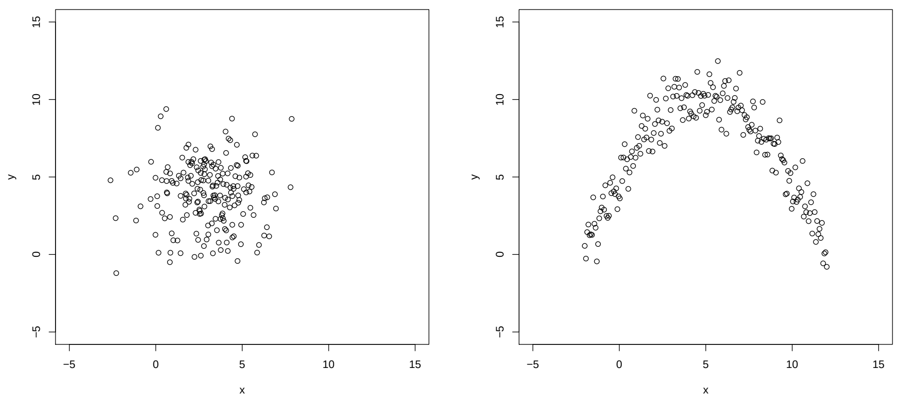
(f) **(2 points)** (i) If the empirical correlation is 1, then the points (x_i, y_i) are perfectly aligned and the corresponding line has a strictly positive slope. Possible data configurations are



(ii) If the empirical correlation is -0.5, then the points (x_i, y_i) form a cloud that has a rough negative trend. Possible data configurations are



(iii) If the empirical correlation is 0, then the points (x_i, y_i) form a cloud that doesn't exhibit a linear trend. Possible data configurations are



(g) (2 points) We have $X_1, \ldots, X_n \sim U(0, \theta)$. The density function of X_i can be written as

$$f(x_i; \theta) = \theta^{-1} \times I(0 < x_i < \theta),$$

where $I(\cdot)$ denotes the indicator function. Then, the likelihood of $x_1, \ldots, x_n$ is

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} \theta^{-1} \times I(0 < x_i < \theta) = \theta^{-n} \times I(0 < x_1, \ldots, x_n < \theta) = \theta^{-n} \times I(m < \theta),$$

where $m = \max(x_1, \ldots, x_n)$. Viewed as a function of θ, L is maximized at $\widehat{\theta} = m$. So $\widehat{\theta} = \max(x_1, \ldots, x_n)$.

(h) (2 points) '(L, U) is a $(1 - \alpha) \times 100\%$ confidence interval for a parameter θ' means that the true value of the unknown parameter θ lies in the random interval (L, U) with probability $(1 - \alpha)$, under repeated sampling from the random experiment that gave rise to (L, U).

Solution 2. (a) (2 points) The marginal probability density function of X is

$$f_X(x) = \int_0^1 f_{X,Y}(x, y)\mathrm{d}y = \int_0^1 c(1+4\theta xy)\mathrm{d}y = c\left[y + 2\theta xy^2\right]_0^1 = c(1+2\theta x), \quad 0 \leq x \leq 1.$$

The integral of $f_X(x)$ is

$$\int_0^1 f_X(x)\mathrm{d}x = \int_0^1 c(1+2\theta x)\mathrm{d}x = c\left[x+\theta x^2\right]_0^1 = c(1+\theta),$$

which must be equal to one for $f_X(x)$ to be a valid density function, so $c = (1+\theta)^{-1}$. Thus,

$$f_X(x) = \frac{1+2\theta x}{1+\theta}, \quad 0 \le x \le 1.$$

By the symmetry of $f_{X,Y}(x,y)$, the marginal probability density function of Y is

$$f_Y(y) = \frac{1+2\theta y}{1+\theta}, \quad 0 \le y \le 1.$$

(b) (2 points) The sunshine quotients on two consecutive days are independent if the random variates X and Y are independent, which is the case if their support is a Cartesian product, and in the support the joint density factorises as $f_{X,Y}(x,y) = f_X(x) \times f_Y(y)$. The support of (X,Y) is $[0,1]\times[0,1]$, which is clearly a Cartesian product. For the joint density to factorize, we must have

$$\frac{1+4\theta xy}{1+\theta} = \frac{1+2\theta(x+y)+4\theta^2 xy}{(1+\theta)^2},$$

or equivalently

$$\theta\{1+4xy+2(x+y)\} = 0,$$

which is satisfied for all (x,y) in $[0,1]\times[0,1]$ only when $\theta = 0$ since the second term on the left-hand side is strictly positive.

[Alternatively we see by inspection that the joint density

$$f_{X,Y}(x,y) = \frac{1+4\theta xy}{1+\theta}, \quad 0 \le x, y \le 1,$$

can factorise into separate functions of x and y only if $\theta = 0$.]

(c) (2 points) The expected sunshine quotient for London is

$$\mathrm{E}(X) = \int_0^1 x \times \tfrac{1+2x}{2}\,\mathrm{d}x = \tfrac{1}{2}\left[\tfrac{x^2}{2} + 2\tfrac{x^3}{3}\right]_0^1 = \tfrac{7}{12}.$$

(d) (2 points) The conditional density of the sunshine quotient tomorrow given today's sunshine quotient is

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1+4\theta xy}{1+2\theta x},$$

which becomes $(1+4y)/3$ for London if $x = 1$.

Solution 3. (a) (2 points) (i) This plot is a Q-Q plot, which is one way to compare a sample $X_1,\ldots,X_n$ with a theoretical distribution F (here the standard normal distribution).

The Q-Q plot displays the sample quantiles

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

plotted against

$$F^{-1}\{1/(n+1)\}, F^{-1}\{2/(n+1)\}, \ldots, F^{-1}\{n/(n+1)\}.$$

The closer the graph is to a straight line, the more the data resemble a sample from F.

(ii) The points in the Q-Q plot are quite well aligned, so the sample looks like a sample from a normal distribution, though with so few observations it is hard to be sure.

(b) (3 points) Based on (a), it seems reasonable to suppose that weighings $X_1, \ldots, X_{10} \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, where μ is the true weight of the suitcase (i.e., the scales are unbiased), and σ^2 is the unknown constant variance. We want to test the null hypothesis $H_0 : \mu = \mu_0 = 20$ against the alternative hypothesis $H_1 : \mu > \mu_0 = 20$.

A $(1 - \alpha)$ confidence interval for μ can be computed as

$$\left(\overline{x} - t_{n-1}(1 - \alpha_L)\tfrac{s}{\sqrt{n}}, \overline{x} - t_{n-1}(\alpha_U)\tfrac{s}{\sqrt{n}}\right),$$

where $t_\nu(p)$ is the p quantile of the Student t distribution with ν degrees of freedom, and $\alpha_L, \alpha_U \geq 0$ are such that $\alpha_L + \alpha_U = \alpha$. Since we want to determine whether the suitcase is overweight, i.e., $\mu > 20$, we set $\alpha_L = \alpha$ and use the one-sided interval (L, ∞). Using the data, we have $\overline{x} = (x_1 + \cdots + x_{10})/10 = 20.55$, and $s = \{9^{-1}\sum_{i=1}^{10}(x_i - \overline{x})\}^{1/2} = 0.495$. A 95% one-sided confidence interval for μ is

$$\left(20.55 - 1.833 \times \tfrac{0.495}{\sqrt{10}}, \infty\right) = (20.263, \infty),$$

which doesn't include 20, so in this case we reject H_0 and conclude that the suitcase is overweight.

(c) (3 points) Here we also want to test the null hypothesis $H_0 : \mu = \mu_0 = 20$ against the alternative hypothesis $H_1 : \mu > \mu_0 = 20$. Let Y_i denote the weight of the ith object. The number of objects being large, we can apply the central limit theorem and approximate the distribution of $\sum_{i=1}^{49} Y_i$ by a normal distribution with known variance. So, we assume that $\sum_{i=1}^{49} Y_i \sim N(\mu, 49 \times 0.004^2)$. A 99% confidence interval is

$$(19.9 - 2.326 \times 7 \times 0.004, \infty) = (19.835, \infty),$$

which includes 20, so in this case we cannot reject H_0; we conclude that Chris is not likely to be fined.

[We also give partial marks for *correct* computations of two-sided confidence limits.]

Solution 4. (a) (2 points) The posterior density of θ given $x_1, \ldots, x_n$ is

$$\begin{aligned}\pi(\theta \mid x_1, \ldots, x_n) &= \frac{f(x_1, \ldots, x_n \mid \theta) \times \pi(\theta)}{f(x_1, \ldots, x_n)} \\ &= \frac{\theta^c(1-\theta)^{n+m-c-1}}{\int_0^1 \theta^c(1-\theta)^{n+m-c-1}\mathrm{d}\theta}, \quad 0 < \theta < 1.\end{aligned}$$

Since

$$\int_0^1 \theta^c(1-\theta)^{n+m-c-1}\mathrm{d}\theta = \frac{\Gamma(c+1)\Gamma(n+m-c)}{\Gamma(n+m+1)},$$

we find that

$$\pi(\theta \mid x_1, \ldots, x_n) = \frac{\Gamma(n+m+1)}{\Gamma(c+1)\Gamma(n+m-c)}\theta^c(1-\theta)^{n+m-c-1}, \quad 0 < \theta < 1.$$

The quantity c represents the number of packets that are corrupted, that is $c = \sum_{i=1}^n x_i$.

(b) (2 points) We simplify the calculation by maximizing the log-posterior density

$$\log \pi(\theta \mid x_1, \ldots, x_n) \propto c \log \theta + (n + m - c - 1)\log(1 - \theta), \quad 0 < \theta < 1,$$

which has a stationary point when

$$\frac{c}{\theta_s} - \frac{n+m-c-1}{1-\theta_s} = 0,$$

or equivalently, when

$$\theta_s = \frac{c}{n+m-1}.$$

Let's check the nature of this stationary point. The curvature of the log-posterior density at $\theta_s > 0$ is

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log \pi(\theta \mid x_1, \ldots, x_n)\bigg|_{\theta=\theta_s} &= -\frac{c}{\theta_s^2} - \frac{n+m-c-1}{(1-\theta_s)^2} \\
&= -\frac{(n+m-1)^2}{c} - \frac{(n+m-1)^2}{n+m-c-1} \\
&= -(n+m-1)^2 \frac{n+m-1}{c(n+m-c-1)} \\
&= -(n+m-1)^2 \frac{1}{c(1-\theta_s)} < 0
\end{aligned}$$

since $c > 0$ and $1 - \theta_s > 0$. So the log-posterior density reaches a maximum at $\theta_s > 0$, and the maximum a posteriori estimate of θ is $\widehat{\theta}_{\mathrm{MAP}} = c/(m+n-1)$.

[Bonus for noticing that it is possible that $c = 0$, in which case the stationary point is at $\theta_s = 0$ but is still a maximum.]

(c) (2 points) The probability that the next packet to arrive will be corrupted, conditional on $x_1, \ldots, x_n$, is

$$\Pr(X_{n+1} = 1 \mid X_1, \ldots, X_n) = \frac{\int_0^1 f(x_1, \ldots, x_{n+1} \mid \theta) \times \pi(\theta)\mathrm{d}\theta}{\int_0^1 f(x_1, \ldots, x_n \mid \theta) \times \pi(\theta)\mathrm{d}\theta}.$$

The numerator is

$$\begin{aligned}
\int_0^1 f(x_1, \ldots, x_{n+1} \mid \theta) \times \pi(\theta)\mathrm{d}\theta &= \int_0^1 \theta^{c+1}(1-\theta)^{n+1-(c+1)} \times (1-\theta)^{m-1}\mathrm{d}\theta \\
&= \int_0^1 \theta^{c+1}(1-\theta)^{n+m-c-1}\mathrm{d}\theta \\
&= \frac{\Gamma(c+2)\Gamma(n+m-c)}{\Gamma(n+m+2)},
\end{aligned}$$

and the denominator is

$$\begin{aligned}
\int_0^1 f(x_1, \ldots, x_n \mid \theta) \times \pi(\theta)\mathrm{d}\theta &= \int_0^1 \theta^{c}(1-\theta)^{n+m-c-1}\mathrm{d}\theta \\
&= \frac{\Gamma(c+1)\Gamma(n+m-c)}{\Gamma(n+m+1)},
\end{aligned}$$

yielding

$$\begin{aligned}
\Pr(X_{n+1} = 1 \mid X_1, \ldots, X_n) &= \frac{\Gamma(c+2)\Gamma(n+m-c)}{\Gamma(n+m+2)} \frac{\Gamma(n+m+1)}{\Gamma(c+1)\Gamma(n+m-c)} \\
&= \frac{c+1}{n+m+1}.
\end{aligned}$$