Ex. 1 **a)** [*4 marks*] The probability space is $(\Omega, \mathcal{F}, \Pr)$, where

$$\Omega = \left\{\omega = (\omega_1, \ldots, \omega_6) \in \mathbb{N}^6 : i, \omega_i \in \{1, \ldots, 6\}\right\},$$

$\mathcal{F}$ is the set of all possible unions and intersections of $\{(\omega_1, \ldots, \omega_6)\}$, and $\Pr(\omega \in \Omega) = 1/6^6$. Let A and B denote the event 'all top faces show the same number' and 'top faces show all different numbers' respectively. Then,

$$\Pr(A) = \frac{6}{6^6} = \frac{1}{6^5}, \quad \Pr(B) = \frac{6!}{6^6} = \frac{5!}{6^5}.$$

Thus, B is (much) more likely than A.

b) [*2 marks*] For all $y \geqslant 0$, we have $Y \leqslant y \iff -\sqrt{y} \leqslant X \leqslant \sqrt{y}$. As $X \geqslant 0$, we have $\Pr(Y \leqslant y) = \Pr(X \leqslant \sqrt{y}) \iff F_Y(y) = F_X(\sqrt{y})$. Therefore,

$$f_Y(y) = \frac{1}{2\sqrt{y}}\, f_X(\sqrt{y}) = \begin{cases} \dfrac{1}{2\sqrt{y}}\lambda \exp(-\lambda\sqrt{y}), & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

c) [*2 marks*] Using independence of X and Y and the law of total probability,

$$\begin{aligned} \Pr(Y = X) &= \sum_{x=1}^{\infty} \Pr(X = x)\Pr(Y = x) \\ &= \sum_{x=1}^{\infty} \{p(1-p)^{x-1}\}^2 \\ &= p^2 \sum_{x=1}^{\infty} \{(1-p)^2\}^{x-1} \\ &= \frac{p^2}{1-(1-p)^2} \\ &= \frac{p}{2-p}. \end{aligned}$$

d) [*2 marks*] Using independence of Z_1 and Z_2, we have

$$\operatorname{cov}(X, Y) = be \operatorname{var}(Z_1) + bf \operatorname{cov}(Z_1, Z_2) + ce \operatorname{cov}(Z_2, Z_1) + cf \operatorname{var}(Z_2) = be + cf.$$

Similarly,

$$\operatorname{var}(X) = \operatorname{cov}(X, X) = b^2 + c^2, \quad \operatorname{var}(Y) = \operatorname{cov}(Y, Y) = e^2 + f^2.$$

So,

$$\operatorname{corr}(X, Y) = \frac{be + cf}{\sqrt{b^2 + c^2} \cdot \sqrt{e^2 + f^2}}.$$

e) [*2 marks*] The marginal density of U is

$$f_U(u) = \int_0^1 cu^2 v \, \mathrm{d}v = \frac{1}{2}cu^2, \quad 0 < u < 2,$$

so

$$f_{V|U=1}(v) = \frac{f_{U,V}(1,v)}{f_U(1)} = \begin{cases} 2v, & 0 < v < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Or, since the support of (U, V) is a rectangle and the density function factorises, the variables are independent, so the density of V is proportional to v, and therefore $f_{V|U=1}(v) = 2vI(0 < v < 1)$; here one would need to check that this integrates to unity.

f) [*1 marks*] Using the continuity of X and the definition of the quantile, we have

$$\Pr(X > x_p) = 1 - \Pr(X \leq x_p) = 1 - \Pr(X \leq x_p) = 1 - p.$$

g) [*2 marks*] It means that the sample median is less influenced by extreme values and outliers than the sample average. This come from the fact that the sample median is the value that separates the data into two groups of equal sample size, when they are sorted in increasing order, whereas the sample average is computed from the values of the observations. For example, if we have $x_1 = 0$, $x_2 = 0.5$, $x_3 = 2$, $x_4 = 5$ and $x_5 = 100$, the sample median is $x_{[0.5]} = 2$ and the sample mean is $\overline{x} = 21.5$. In this example, $\overline{x}$ is more impacted by 100 than $x_{[0.5]}$, which is more consistent with the data that it summarises.

Bonus points for stating the breakdown points of these respective estimators, i.e., 50% and 0%.

h) [*2 marks*] Using the delta method for $g(x) = 1/x$, we have $g'(x) = -1/x^2$ so $1/\overline{\boldsymbol{X}} \dot{\sim} \mathcal{N}\{1/\mu, \sigma^2/(n\mu^4)\}$.

i) [*3 marks*] Using independence of the I_k, we have

$$\mathrm{E}(T) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{E}(I_{2j-1}) \{1 - \mathrm{E}(I_{2j})\} = \frac{1}{n}\sum_{j=1}^{n} p(1-p) = p(1-p).$$

Let $\theta = p(1 - p)$, so $\mathrm{E}(T) = \theta$. Since T is unbiased for θ, we only need $\mathrm{var}(T)$, and this equals $\mathrm{var}\{I_1(1 - I_2)\}/n$. Now since the variables here are Bernoulli,

$$\mathrm{E}[\{I_1(1 - I_2)\}^2] = \mathrm{E}\{I_1(1 - I_2)\} = p(1 - p),$$

so

$$\mathrm{MSE}(T) = \mathrm{var}(T) = n^{-1}\{p(1-p) - p^2(1-p)^2\} = n^{-1}p(1-p)\{1 - p(1-p)\}.$$

Ex. 2 Let $F = 1$ if the transaction is fraudulent and let $F = 0$ otherwise, and let X denote the transaction. Under this model the unconditional mean for a fraudulent payment, $\mathrm{E}(X \mid F = 1) = \infty$, so the bank is wise to impose an upper limit on payments.

a) [*2 marks*] The law of total probability gives

$$\begin{aligned}
\Pr(X > 1) &= \Pr(X > 1 \mid F = 1)\Pr(F = 1) + \Pr(X > 1 \mid F = 0)\Pr(F = 0)\\
&= \int_1^\infty \frac{9}{(1+9x)^2}\,\mathrm{d}x \times 0.01 + 0 \times 0.99\\
&= 0.01\left[-\frac{1}{1+9x}\right]_1^\infty\\
&= \frac{0.01}{10} = 0.001.
\end{aligned}$$

b) [*3 marks*] Bayes' theorem gives

$$\begin{aligned}
\Pr(F = 1 \mid X = x) &= \frac{f_X(x \mid F = 1)\Pr(F = 1)}{f_X(x \mid F = 1)\Pr(F = 1) + f_X(x \mid F = 0)\Pr(F = 0)}\\
&= \frac{0.01 \times 9/(1+9x)^2}{0.01 \times 9/(1+9x)^2 + 0.99 \times 1}, \quad x < 1,
\end{aligned}$$

and obviously $\Pr(F = 1 \mid X = x) = 1$ if $x > 1$. Hence

$$\Pr(F \mid X = x) = \begin{cases} \dfrac{9}{9 + 99(1+9x)^2}, & 0 < x \leqslant 1,\\ 1, & \text{otherwise.} \end{cases}$$

This might seem a bit surprising, because this probability decreases with x, but this is because the density for fraudulent transactions is highest at $x = 0$, which is not very plausible as a model.

c) [*3 marks*] Here we deal only with fraudulent transactions, i.e., $F = 1$, and so we want

$$\mathrm{E}(X \mid F = 1, X \leq 1),$$

since any transaction with $X > 1$ is refused (and thus the bank will not have to pay out). Now

$$\Pr(X \leq 1 \mid F = 1) = \int_0^1 \frac{9}{(1+9x)^2}\,\mathrm{d}x = \left[-\frac{1}{1+9x}\right]_0^1 = 0.9,$$

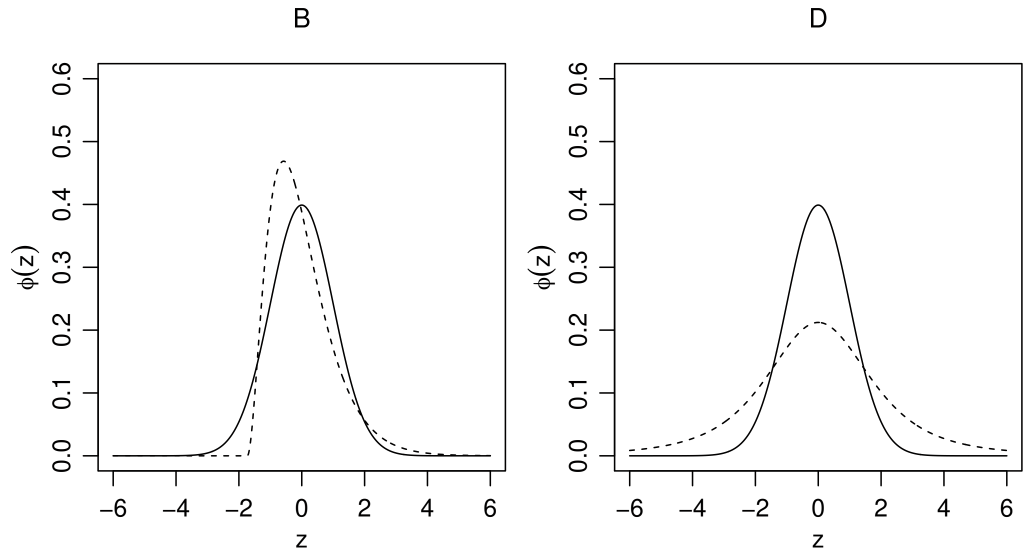
so

$$\begin{aligned}
\mathrm{E}(X \mid F = 1, X \leq 1) &= \frac{1}{0.9}\int_0^1 x \times \frac{9}{(1+9x)^2}\,\mathrm{d}x\\
&= \frac{1}{0.9}\int_0^1 \left\{\frac{1+9x}{(1+9x)^2} - \frac{1}{(1+9x)^2}\right\}\mathrm{d}x\\
&= \frac{1}{9 \times 0.9}\left[\log(1+9x) + \frac{1}{(1+9x)}\right]_0^1\\
&= \frac{1}{9 \times 0.9}\left[\log 10 + \frac{1}{10} - \frac{1}{1}\right]\\
&= \frac{10}{81}(\log 10 - 9/10)\\
&\approx 0.173.
\end{aligned}$$

Ex. 3 **a)** [*2 marks*] $(n-1)s^2 = \sum(x_j - \overline{x})^2 = 0$ implies that $x_j = \overline{x}$ for all j, i.e., all the data are equal. None of the other statements is necessarily true.

b) [*2 marks*] The empirical correlation measures LINEAR association between x and y and always lies in the interval $[-1, 1]$. It does not depend on the units, and if $(x_1, y_1) = (1, -1)$ and $(x_2, y_2) = (1, 1)$ then we could have $r = 0$.

c) [*4 marks*] The data in A are clearly skewed to the right relative to the normal sample, and those in C clearly have heavier tails. The corresponding densities are here :



Ex. 4 **a)** [*2 marks*] (i) The hypergeometric is a discrete distribution, but has limits on its values, so would not be a reasonable model. (ii) The binomial is discrete but implies an upper bound to the score, so this is not a good choice. (iii) The exponential is continuous, so would be inappropriate. (iv) The Poisson is discrete with no upper limit, so seems best.

b) [*4 marks*] The log-likelihood for λ is

$$\ell(\lambda; y) = \sum_{i=1}^{n} \log \Pr(Y_i = y_i; \lambda) = \sum_{i=1}^{n} \log \frac{\lambda^{y_i} \exp(-\lambda)}{y_i!} = \sum_{i=1}^{n} y_i \log \lambda - n\lambda + \text{constant}, \quad \lambda > 0.$$

Taking the derivatives with respect to λ yields

$$\begin{aligned} \ell'(\lambda; y) &= -n + \sum_{i=1}^{n} y_i/\lambda \Rightarrow \widehat{\lambda} = \overline{y}. \\ \ell''(\lambda; y)\big|_{\lambda=\widehat{\lambda}} &= -\sum_{i=1}^{n} y_i/\widehat{\lambda}^2 = -n/\widehat{\lambda} < 0. \end{aligned}$$

So $\widehat{\lambda}$ is indeed a maximum.

In this case the total number of matches is 66 and the total number of goals is 164, so $\widehat{\lambda} = 164/66 \doteq 2.48$.

Using asymptotic normality of the MLE, the 95% confidence interval has limits

$$\widehat{\lambda} \pm z_{0.975}\sqrt{-1/\ell''(\widehat{\lambda}; y)} = \overline{y} \pm 1.96\sqrt{\overline{y}/n} = 2.48 \pm 1.96\sqrt{2.48/66} = [2.10, 2.86].$$

A 95% confidence interval for the mean number of goals scored by FC Sion in the rounds of this competition where it plays is roughly $(2.1, 2.9)$. with mean roughly 2.5.

c) [*2 marks*] The p-value measures the evidence against the null hypothesis, with small values indicating stronger evidence. The given P-value of 3.4×10^{-6} is very strong evidence against the hypothesis of constant λ, and indeed a look at the data suggests that the mean number of goals scored in the first round is roughly 4, but is closer to 2 in subsequent rounds.

d) [*2 marks*] The analysis in b) seems unrealistic in assuming a constant number of goals in each round, since Sion will probably meet better opponents in the later rounds. Analysis in c) confirms that this assumption is very implausible. Constant λ also means that FC Sion plays on average at the same level, which suggests having the same team members, being in the same physical condition, etc., and all these conditions being stable through 15 years. A more realistic assumption would be that λ is lower in higher rounds, inspired by the idea that when a team advances in the competition it will tend to face stronger teams, so its mean scores decrease. One might also allow λ to vary from year to year.