Anomaly Detection:
Assessed coursework (solutions)

Patrick Rubin-Delanchy
patrick.rubin-delanchy@bristol.ac.uk

School of Mathematics, University of Bristol

This sheet gives the solutions to Parts 1 and 2. The code for Part 3 is available in a separate release.

1. **[10 marks]**

   (a) Suppose $X \leq_{st} Y$. Prove that if f is a decreasing function, then

   $$f(X) \geq_{st} f(Y).$$

   Solution. For $a \in \mathbb{R}$,

   $$\begin{aligned} \mathbb{P}\{f(X) \leq a\} &= \mathbb{P}\{X \geq f^{-1}(a)\} \\ &\leq \mathbb{P}\{Y \geq f^{-1}(a)\} \\ &= \mathbb{P}\{f(Y) \leq a\}. \end{aligned}$$

   (b) If $X_1, \dots, X_n$ are independent, $Y_1, \dots, Y_n$ are independent, and for each i we have $X_i \leq_{st} Y_i$, then it is known that:

   $$\sum_{i=1}^n X_i \leq_{st} \sum_{i=1}^n Y_i.$$

   Using this result, together with the result of 1(a), prove that

   $$-2\sum_{i=1}^n \log(U_i) \geq_{st} -2\sum_{i=1}^n \log(P_i),$$

   if $U_1, \dots, U_n$ are independent uniform random variables on $[0,1]$, and $P_1, \dots, P_n$ are independent random variables with stochastically larger than uniform distributions.
   Solution. $-2\log(\cdot)$ is a decreasing function. Letting $Y_i = -2\log(U_i)$ and $X_i = -2\log(P_i)$, we have $X_i \leq_{st} Y_i$ using 1(a), and therefore

   $$-2\sum_{i=1}^n \log(P_i) = \sum_{i=1}^n X_i \leq_{st} \sum_{i=1}^n Y_i = -2\sum_{i=1}^n \log(U_i).$$

(c) In the context of combining p-values, we will say that a procedure is conservative if the probability of the combined p-value being lower than α is lower than α. Based on the above, and justifying your answer, which of the following is true?
Applying Fisher's method with discrete p-values as input gives:

i. a conservative procedure.
ii. a liberal procedure.
iii. neither.

Solution. If the observed p-values are $p_1, \ldots, p_n$, then the combined p-value using Fisher's method is

$$S_{\chi^2_{2n}} \left\{ -2 \sum_{i=1}^{n} \log(p_i) \right\},$$

where $S_{\chi^2_k}(a) = \mathbb{P}(X \geq a)$ is the survival function of a chi-squared random variable with k degrees of freedom. Under the null hypothesis we have:

$$\begin{aligned}
\mathbb{P}_0 \left[S_{\chi^2_{2n}} \left\{ -2 \sum_{i=1}^{n} \log(P_i) \right\} \leq \alpha \right] &= \mathbb{P}_0 \left\{ -2 \sum_{i=1}^{n} \log(P_i) \geq S^{-1}_{\chi^2_{2n}}(\alpha) \right\} \\
&\leq \mathbb{P}_0 \left\{ -2 \sum_{i=1}^{n} \log(U_i) \geq S^{-1}_{\chi^2_{2n}}(\alpha) \right\} \\
&= S_{\chi^2_{2n}} \left\{ S^{-1}_{\chi^2_{2n}}(\alpha) \right\} = \alpha,
\end{aligned}$$

reversing the inequality in the second line because $S_{\chi^2_{2n}}$ is decreasing. The procedure is therefore conservative.

2. **[10 marks]** Assume the following "hierarchical" hurdle model for a communication network involving n entities. The number of communications at time t from entity i to entity j (allowing $i = j$, e.g. sending an email to oneself) is modelled as

$$\mathrm{d}N_{ij}(t) = \mathrm{d}O_i(t)\mathrm{d}A_{ij}(t)\{\mathrm{d}B_{ij}(t) + 1\},$$

where $O_i(t)$, $A_{ij}(t)$ and $B_{ij}(t)$ are discrete time counting processes with $\mathrm{d}O_i(t) \in \{0, 1\}, \mathrm{d}A_{ij}(t) \in \{0, 1\}$, $\mathrm{d}B_{ij}(t) \geq 0$. This model could be relevant, for example, in a context where i only sends messages when "online", i.e. $\mathrm{d}O_i(t) = 1$. Assume that:

(a) $\mathrm{d}O_i(t)$ is a sequence of independent and identically distributed Bernoulli variables with success probability μ_i.

(b) Conditional on $\mathrm{d}O_i(t) = 1$, we have $\sum_{j=1}^{n} \mathrm{d}A_{ij}(t) \geq 1$ with probability one (i sends at least one message when online).

(c) $\mathrm{d}B_{ij}(t)$ is a sequence of independent and identically distributed Poisson variables with mean λ_{ij}.

Prove that:

(a) Under a Beta(α_O, β_O) prior for μ_i we have,

$$[\mu_i \mid N_{ij}(\tau), j = 1, \dots, n, \tau = 1, \dots, t] \sim \mathrm{Beta}\{\alpha_O + O_i(t), \beta_O + t - O_i(t)\}.$$

Solution. We have:

$$\begin{aligned} p\{\mu_i \mid N_{ij}(\tau), j = 1, \dots, n, \tau = 1, \dots, t\} &= p\{\mu_i \mid O_i(\tau), \tau = 1, \dots, t\} \\ &\propto p\{\mathrm{d}O_i(1), \dots, \mathrm{d}O_i(t) \mid \mu_i\} p(\mu_i) \\ &\propto \mu_i^{O_i(t)} (1 - \mu_i)^{t - O_i(t)} \mu_i^{\alpha_O - 1} (1 - \mu_i)^{\beta_O - 1} \\ &\propto \mu_i^{\{O_i(t) + \alpha_O\} - 1} (1 - \mu_i)^{\{t - O_i(t) + \beta_O\} - 1}, \end{aligned}$$

from which we recognise the functional form of a Beta$\{\alpha_O + O_i(t), \beta_O + t - O_i(t)\}$ distribution. (**5 marks**)

(b) Under a Gamma(α_B, β_B) prior for λ_{ij} we have,

$$[\lambda_{ij} \mid N_{ij}(\tau), \tau = 1, \dots, t] \sim \mathrm{Gamma}\{\alpha_B + N_{ij}(t) - A^*_{ij}(t), \beta_B + A^*_{ij}(t)\},$$

where

$$A^*_{ij}(t) = \sum_{\tau=1}^{t} \mathbb{I}\left(\mathrm{d}N_{ij}(\tau) \geq 1\right).$$

Solution. We have:

$$\begin{aligned} p\{\lambda_{ij} \mid N_{ij}(\tau), \tau = 1, \dots, t\} &= p\{\lambda_{ij} \mid (\mathrm{d}A_{ij}(\tau) : \mathrm{d}O_i(\tau) = 1, \tau = 1, \dots, t)\} \\ &\propto p\{(\mathrm{d}A_{ij}(\tau) : \mathrm{d}O_i(\tau) = 1, \tau = 1, \dots, t) \mid \lambda_{ij}\} p(\lambda_{ij}) \\ &\propto \lambda_{ij}^{N_{ij}(t) - A^*_{ij}(t)} \, \mathrm{e}^{-A^*_{ij}(t)\lambda_{ij}} \, p(\lambda_{ij}) \\ &\propto \lambda_{ij}^{\alpha_B + N_{ij}(t) - A^*_{ij}(t) - 1} \, \mathrm{e}^{-\{\beta_B + A^*_{ij}(t)\}\lambda_{ij}}, \end{aligned}$$

from which we recognise the functional form of a Gamma$\{\alpha_B + N_{ij}(t) - A^*_{ij}(t), \beta_B + A^*_{ij}(t)\}$ distribution.