Statistical Theory: Exercise Sheet 1 — Corrections

September 25, 2018; 8:15-10:00 in MA 12.

Tomáš Rubín, tomas.rubin@epfl.ch

Exercise 1. If X is exponentially distributed with intensity λ, what is the distribution of $Y = \lfloor X \rfloor$? ($\lfloor x \rfloor$ stands for the integer part of x.)

Solution to Exercise 1. Y has a discrete distribution (it takes only values $0, 1, 2, \dots$). For $k = 0, 1, \dots$ calculate

$$P(Y = k) = P(\lfloor X \rfloor = k) = P(X \in [k, k+1)) = \int_k^{k+1} \lambda e^{-\lambda x} dx = e^{-k\lambda}(1 - e^{-\lambda})$$

Y is geometric with parameter (success probability) $1 - e^{-\lambda}$.

Exercise 2. Suppose that $S \sim \text{Exp}(\lambda)$, $C \sim \text{Exp}(\gamma)$ are independent. Define $T = \min(S, C)$ and $D = 1[T = S]$. Find the joint ditribution of T and D and their marginal distributions. Are T and D independent?

Solution to Exercise 2. T has a continuous distribution D has a discrete distribution with values 0 and 1. We can specify the join distribution by computing the joint distribution function. Alternatively, since D is discrete, we can specify the distribution by calculating the probabilities below.

Compute for $t \geq 0$

$$P(T \leq t, D = 1) = P(S \leq t, S \leq C) = \int_0^t \left(\int_s^\infty \lambda e^{-\lambda s} \gamma e^{-\gamma c} dc \right) ds = \frac{\lambda}{\lambda + \gamma}(1 - e^{-(\lambda+\gamma)t})$$

and similarly

$$P(T \leq t, D = 0) = P(C \leq t, C \leq S) = \int_0^t \left(\int_c^\infty \lambda e^{-\lambda s} \gamma e^{-\gamma c} ds \right) dc = \frac{\gamma}{\lambda + \gamma}(1 - e^{-(\lambda+\gamma)t}).$$

The marginal distribution of D is Bernoulli with parameter $\lambda/(\lambda + \gamma)$ (compute the limit for $t \to \infty$), T is marginally $\text{Exp}(\lambda + \gamma)$ (sum the above equations). They are independent because the product of the marginal distribution function is the joint distribution function.

Exercise 3. Consider a random vector $(X, Z)^\mathsf{T}$. Let the marginal distribution of Z be exponential with parameter γ, i.e., with density

$$f_Z(z) = \gamma e^{-\gamma z} 1_{(0,\infty)}(z).$$

Suppose that the conditional distribution of X given $Z = z$ is Poisson with parameter λz, that is,

$$f_{X|Z}(x|z) = P(X = x|Z = z) = \begin{cases} \frac{(\lambda z)^x}{x!} e^{-\lambda z}, & x = 0, 1, \ldots, \quad z > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the joint density $f_{X,Z}(x, z)$ (for all (x, z)).

(b) Compute the marginal (unconditional) distribution of X. Which known distribution is it?

(c) Find $\mathsf{E}[X|Z]$.

(d) Find $\mathsf{E}\,X$ using the formula $\mathsf{E}\,X = \mathsf{E}\{\mathsf{E}[X|Z]\}$.

(e) Find $\mathrm{var}[X|Z]$.

(f) Compute $\mathrm{var}\,X$ using the formula

$$\mathrm{var}\,X = \mathsf{E}\{\mathrm{var}[X|Z]\} + \mathrm{var}\{\mathsf{E}[X|Z]\}.$$

(g) Compute the conditional density $f_{Z|X}(z|x)$. Which known distribution is it?

Remark: The variable X can be interpreted as the value of a Poisson process $N(t)$ with intensity λ at random $\mathrm{Exp}(\gamma)$-distributed time $t = Z$ (independent of $N(\cdot)$), that is, $X \sim N(Z)$; also, X follows a Poisson model with an exponential frailty (random effect).

Solution to Exercise 3.

(a) $f_{X,Z}(x, z) = \gamma e^{-(\lambda+\gamma)z} \frac{(\lambda z)^x}{x!} 1_{\mathbb{N}_0 \times \mathbb{R}^+}(x, z)$.

(Use the formula $f_{X|Z}(x|z) = \frac{f_{X,Z}(x,z)}{f_Z(z)}$ and be careful about the values the random variables can take.)

(b) X is geometric with parameter (success probability) $\gamma/(\lambda + \gamma)$.

(From the join density "integrate out" the z variable: $f_X(x) = \int_0^\infty f_{X,Z}(x, z)dz$. You might need to use the gamma function to calculate the integral.)

(c) $\mathsf{E}[X|Z] = \lambda Z$.

(Conditional distribution of X conditioned on $Z = z$ is Poisson with parameter λz. This distribution has mean λz. Substitute z for Z and get $\mathsf{E}[X|Z] = \lambda Z$.)

(d) $\mathsf{E}\,X = \mathsf{E}(\lambda Z) = \lambda/\gamma$.

(Either calculate $\mathsf{E}\,X$ the hard-way using the marginal probability function (frequencies) from (b) or use (c) and realise that $\mathsf{E}\,X = \mathsf{E}[\mathsf{E}[X|Z]] = \mathsf{E}[\lambda Z] = \lambda/\gamma$.)

(e) $\mathrm{var}[X|Z] = \lambda Z$.

(Similarly as in (c), the variance o Poisson distribution with parameter λz is λz.)

(f) $\mathrm{var}\,X = \mathsf{E}(\lambda Z) + \mathrm{var}(\lambda Z) = \lambda/\gamma + \lambda^2/\gamma^2$.

(By the formula from the hint.)

(g) $f_{Z|X}(z|x) = \frac{(\lambda+\gamma)^{x+1}}{x!} e^{-(\lambda+\gamma)z} z^x$ for $z > 0$, $x = 0, 1, \ldots$, $f_{Z|X}(z|x) = 0$ otherwise. Hence $Z|X = x \sim \Gamma(\lambda + \gamma, x + 1)$.

(Use the formula $f_{Z|X}(z|x) = \frac{f_{X,Z}(x,z)}{f_X(x)}$ but be careful about the values the random variables take. Z is continuous and X is discrete.)

Exercise 4. Suppose that the observation X follows the model of the previous exercise. Is this model for X (parametrised by (λ, γ)) identifiable?

Solution to Exercise 4. X is geometric with parameter $p = \gamma/(\lambda + \gamma)$. The mapping $(\lambda, \gamma) \mapsto p = \gamma/(\lambda + \gamma)$ is not one-to-one, hence the model is not identifiable.

In particular, one could easily find $(\lambda_1, \gamma_1) \neq (\lambda_2, \gamma_2)$ such that $\frac{\gamma_1}{\lambda_1+\gamma_1} = \frac{\gamma_2}{\lambda_2+\gamma_2}$.

Exercise 5. Let the observations $Y_1, \ldots, Y_n$ satisfy the regression model

$$Y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i,$$

where x_{ij} are some constants and e_i are independent variables with distribution $\mathrm{N}(0, \sigma^2)$. Give a necessary and sufficient condition on $x_{ij}, i = 1, \ldots, n, j = 1, \ldots, p$ for identifiability of this model (parametrised by $(\beta_1, \ldots, \beta_p, \sigma^2)$).

Solution to Exercise 5. The columns of the $(n \times p)$-matrix $X = (x_{ij})$ must be linearly independent to guarantee that two different linear combinations of them give different vectors of means, hence different distributions.

Exercise 6. Consider a young couple sharing an apartment, and the random variables:

- T_b ...the waiting time until the couple have a baby,
- T_m ...the waiting time until the couple move,
- T_d ...the waiting time until the couple break up,
- T ...the waiting time until the first of the above-mentioned events takes place,
- N ...the number of people living in the apartment at time T:

$$N = \begin{cases} 3 & \text{if } T = T_b, \\ 0 & \text{if } T = T_m, \\ 1 & \text{if } T = T_d. \end{cases}$$

Suppose that the variables T_b, T_m, T_d are independent, and $T_b \sim \mathrm{Exp}(\beta)$, $T_m \sim \mathrm{Exp}(\mu)$, $T_d \sim \mathrm{Exp}(\delta)$.

(a) Find the distribution of T.

(b) Find the joint distribution of T and N.

(c) Find the distribution of N.

(d) Is the model for T parametrized by (β, μ, δ) identifiable? If not, re-parametrize the model (i.e. parametrize it by a function of the three parameters) so that it becomes identifiable.

(e) Is the model for T and N parametrized by (β, μ, δ) identifiable? If not, re-parametrize the model so that it becomes identifiable.

Solution to Exercise 6. (a) $T = \min(T_b, T_m, T_d) \sim \mathrm{Exp}(\beta + \mu + \delta)$.

(b) $T \sim \mathrm{Exp}(\beta + \mu + \delta)$,

$$\mathsf{P}(N = 3) = \frac{\beta}{\beta + \mu + \delta}, \quad \mathsf{P}(N = 0) = \frac{\mu}{\beta + \mu + \delta}, \quad \mathsf{P}(N = 1) = \frac{\delta}{\beta + \mu + \delta},,$$

and T and N are independent.

(c)

$$\mathsf{P}(N = 3) = \frac{\beta}{\beta + \mu + \delta}, \quad \mathsf{P}(N = 0) = \frac{\mu}{\beta + \mu + \delta}, \quad \mathsf{P}(N = 1) = \frac{\delta}{\beta + \mu + \delta},.$$

(d) No. Different vectors (β, μ, δ) can give the same sum $\beta + \mu + \delta$. A model parametrized by $\lambda = \beta + \mu + \delta$ is identifiable.

(e) Yes. Different values of λ give different distributions of T, whereas different values of β/λ, μ/λ, and δ/λ give different distributions of N.