EXAMPLE: INFERENCE FOR POISSON SAMPLING

ERIK QUAEGHEBEUR AND GERT DE COOMAN

1. Representing the data

Observed numbers of emails arriving in Gert's mailbox between 9am and 10am on ten consecutive Mondays:

$$2 \quad 7 \quad 5 \quad 3 \quad 3 \quad 3 \quad 1 \quad 5 \quad 1 \quad 2.$$

Their mean—*the sample mean*—is $\bar{x} = 3.2$. Sorted, we get the *order statistics*:

$$1 \quad 1 \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \quad 5 \quad 5 \quad 7.$$

So the observed values j are

$$1 \quad 2 \quad 3 \quad 5 \quad 7.$$

The observed number of occurrences n_j for each value is:

$$2 \quad 2 \quad 3 \quad 2 \quad 1.$$

In Figures 1 and 2 we give some visualizations of this data.

In an NPI-style fashion, we can associate equal probability mass to the interval between each ordered observation (and the presumed borders of the set of possible observations, 0 and $+\infty$) and shift it around in the interval between its neighbors. This gives, for example, a mass in $[\max\{0, \frac{n_j-1}{N+1}\}, \frac{n_j+1}{N+1}]$ for every value j. The corresponding visualization can be found in Figure 3.

Another simple approach is the linear-vacuous model, which can also be interpreted as adding a number of pseudo-observations that can fall anywhere. Examples of this are visualized in Figures 4 and 5.

2. Bringing in the sampling model

Up until now, we have assumed nothing about the distribution of the number of emails arriving in Gert's mailbox between 9am and 10am on Mondays. A typical probabilistic model for the number of events occurring in a fixed interval Δt (of, e.g., time) is the Poisson distribution. Underlying this distribution is the assumption that these events occur with a known average rate $\dot{\lambda}$ and independently of the time since the last event.

The probability mass assigned by the Poisson distribution with real-valued parameter $\lambda := \dot{\lambda}\Delta t \geq 0$ to any natural number $z \geq 0$ is

$$p_{\mathrm{Ps}}(z\,|\,\lambda) := \frac{\lambda^z \exp(-\lambda)}{z!}.$$

The expected number of events in this interval is equal to λ: the rate $\dot{\lambda}$ times the length Δ of the interval. Let us use our sample mean, i.e., set $\lambda := 3.2$. The resulting model is visualized in Figure 6.

When reversing the roles of the parameter and values in the expression of the probability mass function p_λ, we get the *likelihood* function of the Poisson process:

$$L_z(\lambda) := \frac{\lambda^z \exp(-\lambda)}{z!}.$$

Date: 24 July 2014.

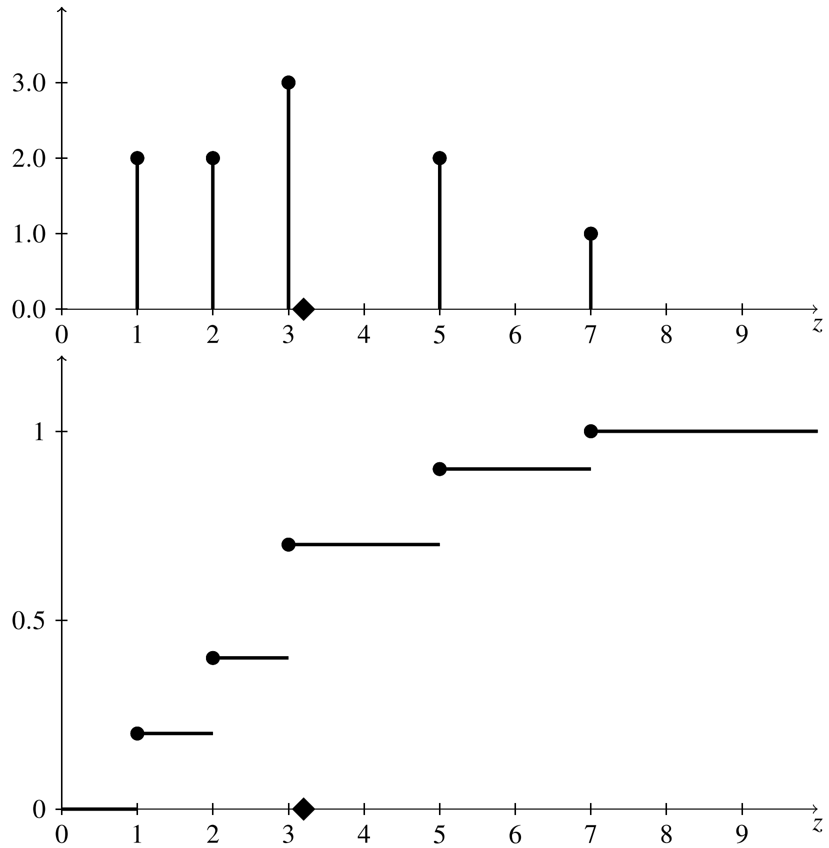

FIGURE 1. Plots for the email arrival numbers data set: on top a comb plot of the occurrences; on the bottom the empirical cumulative distribution function. The sample mean is indicated with a diamond-shaped dot.

The above expression for a single observation can be straightforwardly generalized to one for multiple observations if we assume these observations are *iid*, independently and identically distributed. We get, for a sample x of length n:

$$L_x(\lambda) := p(x\,|\,\lambda) = \prod_{i=1}^{n} p_{\mathrm{Ps}}(x_i\,|\,\lambda) = \prod_{i=1}^{n} L_{x_i}(\lambda) = \frac{\lambda^{n\bar{x}}\exp(-n\lambda)}{\prod_{i=1}^{n} x_i!}.$$

It can be shown that the *maximum likelihood estimate* of λ is equal to the sample mean $\bar{x}$. Fixing n, the sample mean $\bar{x}$ is a sufficient statistic ($L_x(\lambda) = f_n(\lambda,\bar{x})g_n(\lambda)h(x)$). So the last plot is actually the plot of the distribution function of the Poisson distribution with the maximum likelihood as the rate parameter. We can heuristically 'create' an imprecise estimate using confidence intervals. We here consider the 90%-confidence interval and visualize it in Figure 7.

3. BRINGING IN THE PRIOR

Now, in a Bayesian setting, a prior is typically assumed over the set of possible parameter values. So in our example, over the nonnegative reals. Let p be the probability density function corresponding to this prior, then by Bayes's rule (for probability densities) the posterior probability density's expression is

$$p(\lambda\,|\,x) \propto L_x(\lambda)p(\lambda).$$

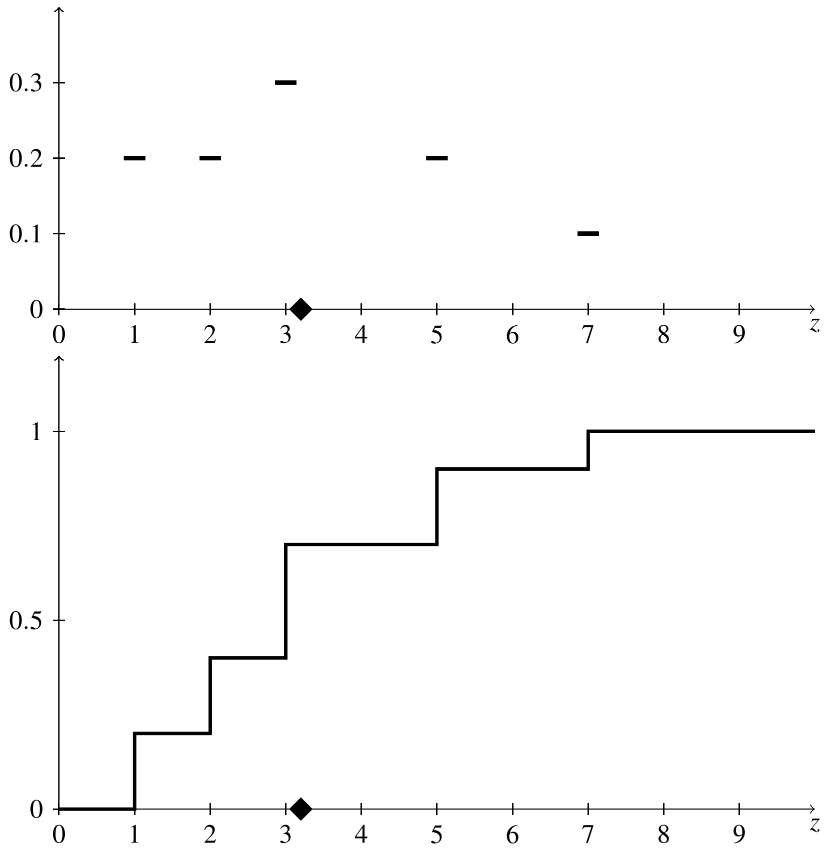

FIGURE 2. Plots for the email arrival numbers data set: on top a plot of the occurrence frequency; on the bottom the empirical cumulative distribution function as a staircase plot.

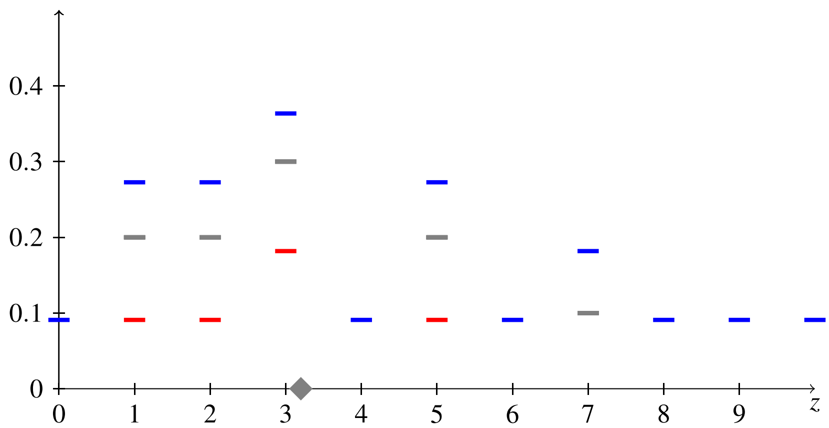

FIGURE 3. Lower and upper probability mass function plot for the NPI model. (The empirical cumulative distribution is given in gray.)

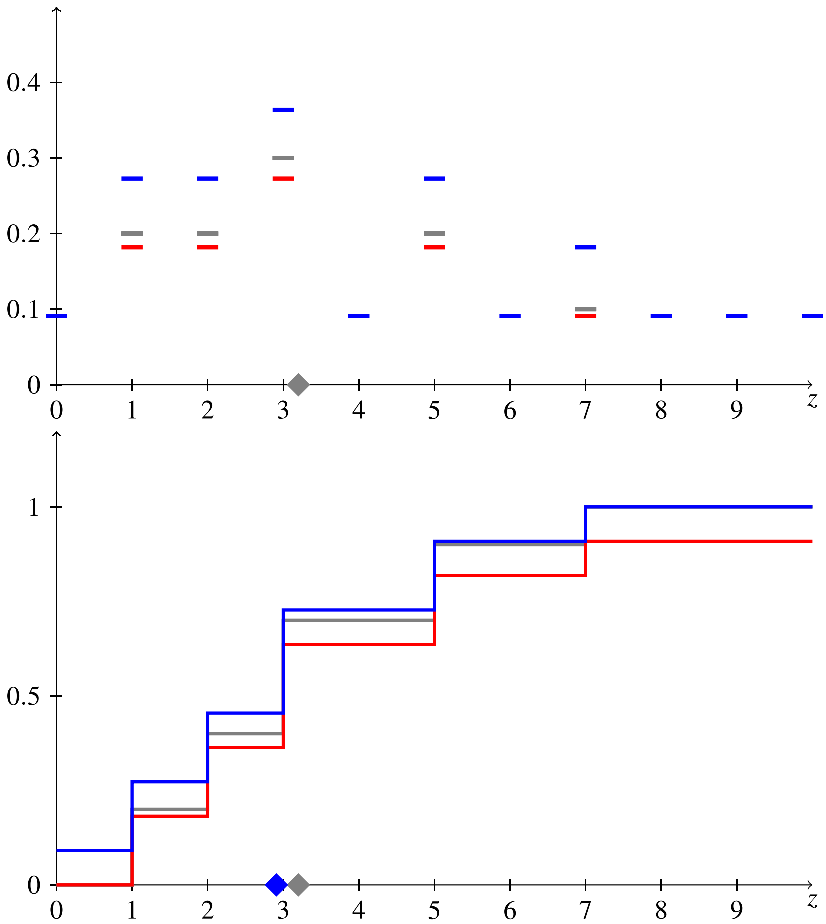

FIGURE 4. Lower and upper probability mass function plot and p-box plot for the linear-vacuous model with $s = 1$.

For general priors the posterior are difficult to obtain in closed analytic form. Therefore, for mathematical convenience, so-called *conjugate* priors are used. The conjugate prior for Poisson sampling is the Gamma distribution, with density

$$p_{\mathrm{Ga}}(\lambda \,|\, \alpha, \beta) := \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda).$$

For later convenience, we will use another parametrization, with $s := \beta$ and $t := \frac{\alpha}{\beta}$, so $st = \alpha$ and

$$p_{\mathrm{Ga}}(\lambda \,|\, s, t) := \frac{s^{st}}{\Gamma(st)} \lambda^{st-1} \exp(-s\lambda),$$

with parameters $s > 0$ and $t > 0$, and Γ the Gamma function. Recall that the (prior) mean is given by $\frac{\alpha}{\beta} = t$ and the (prior) variance by $\frac{\alpha}{\beta^2} = \frac{t}{s}$.

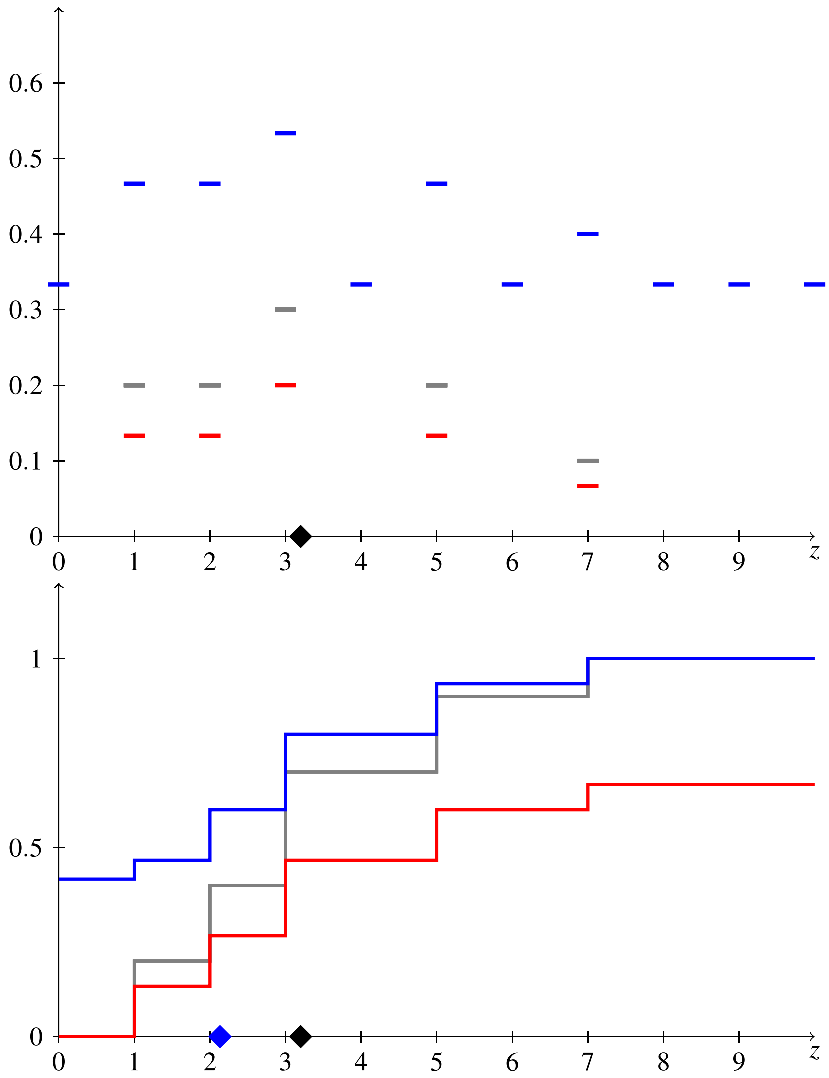

FIGURE 5. Lower and upper probability mass function plot and p-box plot for the linear-vacuous model with $s = 5$.

Due to the compatibility of the expressions of likelihood and conjugate prior, we can directly write down the expression for the posterior:

$$\begin{aligned}
p(\lambda \,|\, x) &\propto L_x(\lambda) p_{\mathrm{Ga}}(\lambda \,|\, s, t) \\
&\propto \lambda^{n\bar{x}} \exp(-n\lambda) \lambda^{st-1} \exp(-s\lambda) \\
&= \lambda^{n\bar{x}+st-1} \exp(-(n+s)\lambda) \\
&\propto p_{\mathrm{Ga}}\left(\lambda \,\middle|\, n+s, \frac{n\bar{x}+st}{n+s}\right).
\end{aligned}$$

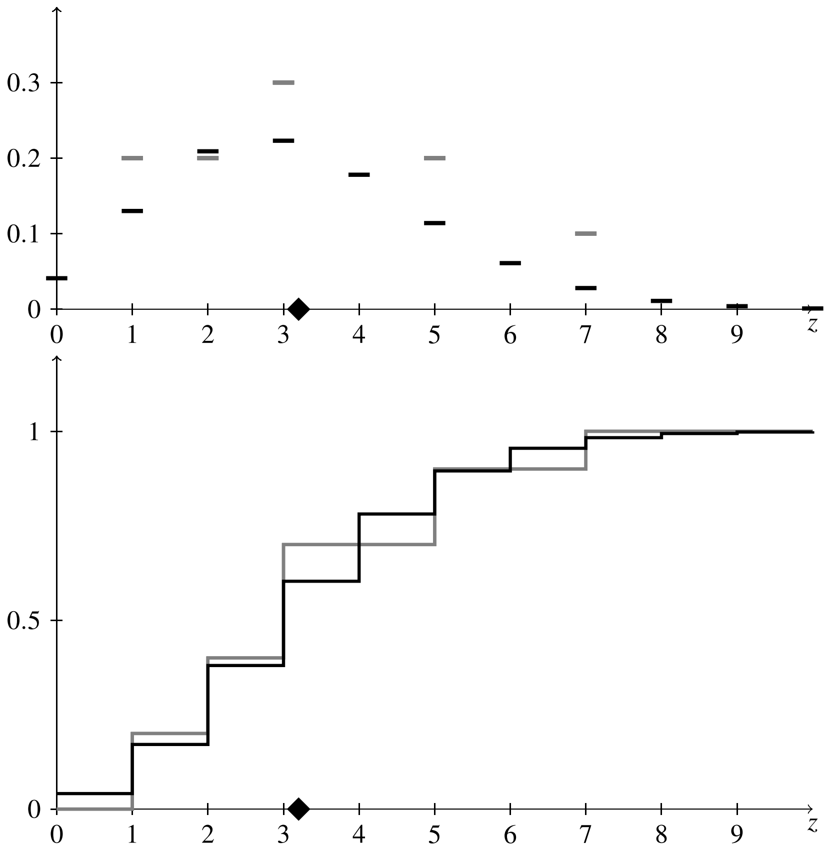

FIGURE 6. Probability mass function plot and cumulative plot for the Poisson model with sample mean value as the parameter value.

So, by normalization, the posterior is of the same conjugate form as the prior, now with parameters $n+s$ and $\frac{n\bar{x}+st}{n+s}$. The latter parameter is also the posterior mean. The posterior mode is $\max\{0, \frac{n\bar{x}+st-1}{n+s}\}$. The posterior variance is $\frac{n\bar{x}+st}{(n+s)^2}$.

When asked, Gert thinks he receives 'about five' emails within the time interval under study, so we ascribe to him a prior with $t := 5$. The precise models we consider are visualized in Figures 8 and 9. We see that the first parameter determines the variance (actually, the prior's variance is $\frac{t}{s}$ and the posterior's variance is $\frac{n\bar{x}+st}{(n+s)^2}$, so the variance also scales with the mean's magnitude). More importantly, by looking at how the mean shifts, we see that the initial value s of the first parameter determines the relative weight of the prior in the posterior; it can be seen as determining the learning rate.

In a second instance, Gert says a more honest assessment of the number of emails he receives within the time interval under study is 'usually somewhere between 2 and 8', so we now ascribe to him a set of priors with $t \in [2,8]$. The imprecise models we consider are visualized in Figures 10 and 11. In the last plot, of sets of prior and posterior densities, only the extreme points are plotted, so the figure may give an incorrect insight if not interpreted correctly.

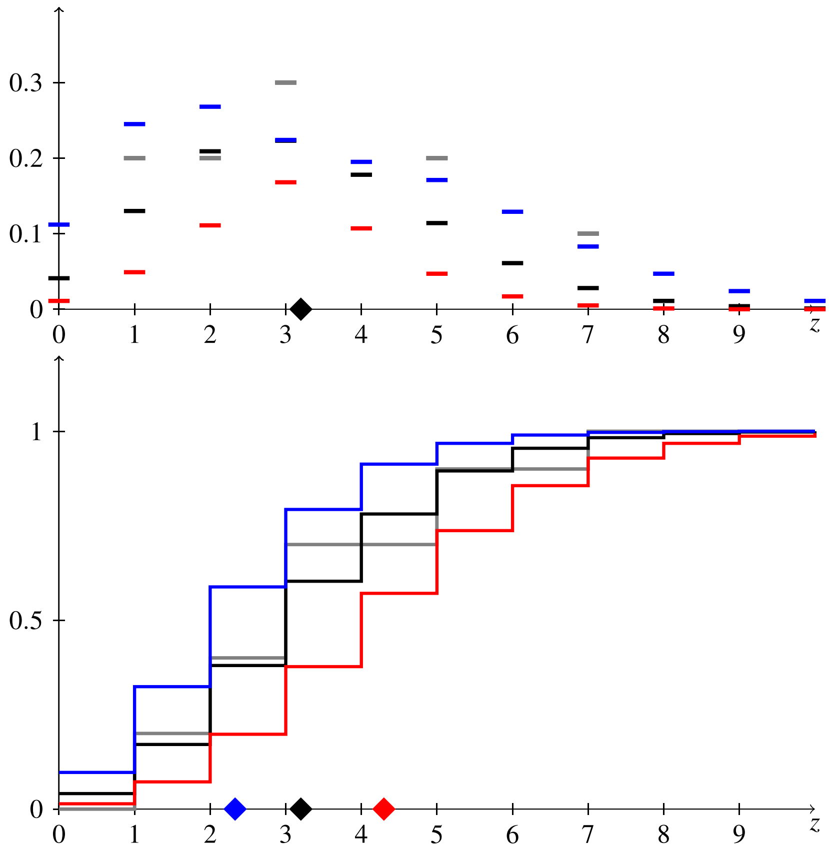

FIGURE 7. Lower and upper probability mass function plot and p-box plot for the heuristic confidence-interval based model.

4. FROM PARAMETRIC TO PREDICTIVE INFERENCE

Up until now in our discussion of the Bayesian approach to inference, we have been looking at parametric inference, i.e., inference about the parameter λ.[1] We can derive predictive inference models from these, i.e., to do inference about a sequence Y of m future observations. We know that if the parameter of the sampling model—in our case a Poisson process with rate λ—is known, that the model for the sequence follows by the iid assumption; to wit, the joint probability mass function with expression

$$p(y\,|\,\lambda) := \prod_{i=1}^{m} p_{\mathrm{Ps}}(y_i\,|\,\lambda).$$

But because we are uncertain about the parameter value, we need to take our model for λ into account. Assume for now that we use a single prior or posterior conjugate distribution for this, then the predictive probability mass function follows by using this conjugate distribution to average over all possible values

[1]Other ways in which the terms parametric and non-parametric are contrasted, is to indicate whether any parameter is involved or some underlying parametric model is used. For example the NPI-inspired inference method is non-parametric in this sense, but its generalization to the arbitrary linear-vacuous case is not. Predictive inference methods based on conjugate parametric priors would also not be called non-parametric in this sense. The take-home lesson is: make sure you know what is meant when somebody makes the distinction parametric vs. non-parametric.

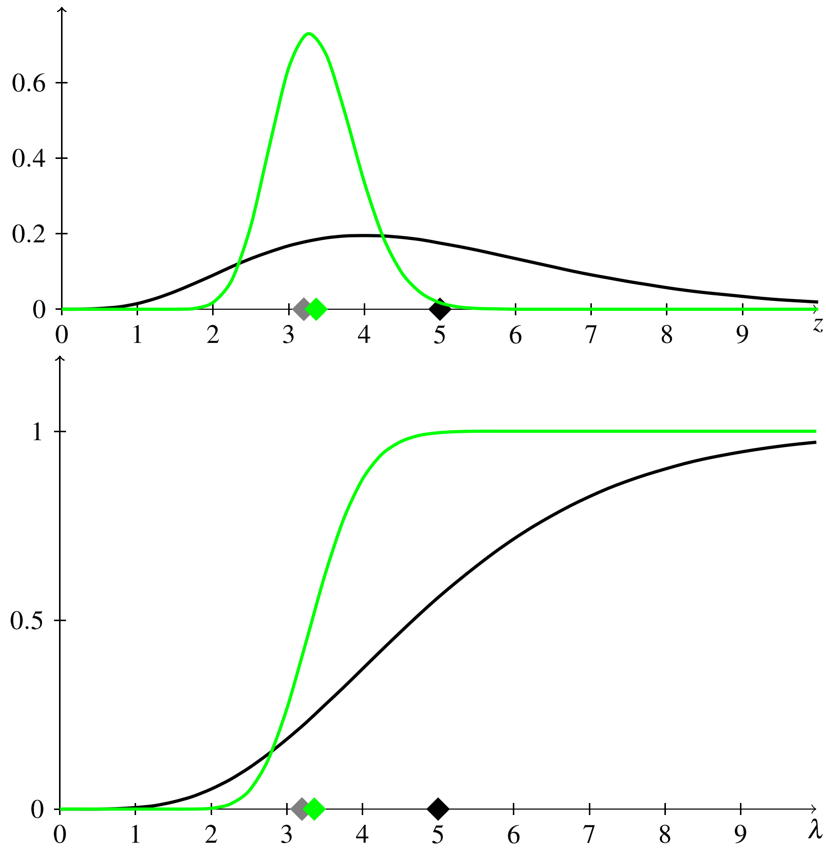

FIGURE 8. Probability density function plot and cumulative distribution function plot for the Gamma model with $s := 1$: prior (in black) and the corresponding posterior (in green).

for and therefore integrate out λ:

$$\begin{aligned}
p(y\,|\,\tilde{s},\tilde{t}) &= \int_0^\infty p(y\,|\,\lambda)p(\lambda\,|\,\tilde{s},\tilde{t})\mathrm{d}\lambda \\
&= \frac{1}{\prod_{i=1}^m y_i!}\frac{\tilde{s}^{\tilde{s}\tilde{t}}}{\Gamma(\tilde{s}\tilde{t})}\frac{\Gamma(m\bar{y}+\tilde{s}\tilde{t})}{(m+\tilde{s})^{m\bar{y}+\tilde{s}\tilde{t}}}\int_0^\infty p\left(\lambda\,\middle|\,m+\tilde{s},\frac{m\bar{y}+\tilde{s}\tilde{t}}{m+\tilde{s}}\right)\mathrm{d}\lambda \\
&= \frac{\Gamma(m\bar{y}+\tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t})\prod_{i=1}^m y_i!}\left(\frac{\tilde{s}}{m+\tilde{s}}\right)^{\tilde{s}\tilde{t}}\frac{1}{(m+\tilde{s})^{m\bar{y}}},
\end{aligned}$$

where $\tilde{s}$ and $\tilde{t}$ refer to either the prior or the posterior parameters and we made use of the fact that $p(y\,|\,\lambda) = L_y(\lambda)$. In general, this expression cannot be simplified further and needs to be evaluated numerically in this form.

We will now focus on immediate prediction, prediction of the next observation Z, so with $m := 1$. Then the predictive probability mass function's expression becomes:

$$p(z\,|\,\tilde{s},\tilde{t}) := \frac{\Gamma(z+\tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t})z!}\left(\frac{\tilde{s}}{1+\tilde{s}}\right)^{\tilde{s}\tilde{t}}\frac{1}{(1+\tilde{s})^z} = \binom{z+\tilde{s}\tilde{t}-1}{z}\left(1-\frac{1}{1+\tilde{s}}\right)^{\tilde{s}\tilde{t}}\left(\frac{1}{1+\tilde{s}}\right)^z,$$

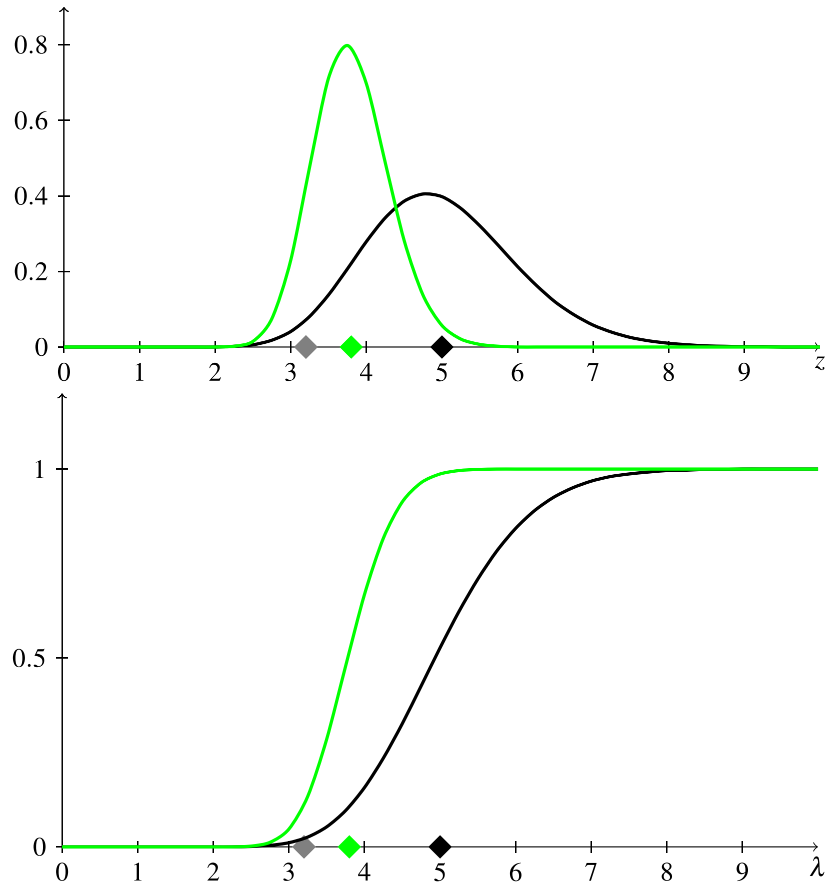

FIGURE 9. Probability density function plot and cumulative distribution function plot for the Gamma model with $s := 5$: prior (in black) and the corresponding posterior (in green).

where the first factor is a generalized binomial coefficient. This is the expression of the negative binomial distribution's probability mass function, with $\frac{1}{1+\tilde{s}}$ the probability of 'success' and $\tilde{s}\tilde{t}$ the possibly non-integer number of 'failures' that determine when sampling is stopped. (It is unclear whether this can be insightfully linked to the email arrival times context we sketched.) Its mean is $\tilde{t}$. In Figure 12 and 13, we give visualizations of the immediate predictive models corresponding to the parametric models given above. In Figure 14 and 15, we give visualizations of the imprecise immediate predictive models corresponding to the imprecise parametric models given above.

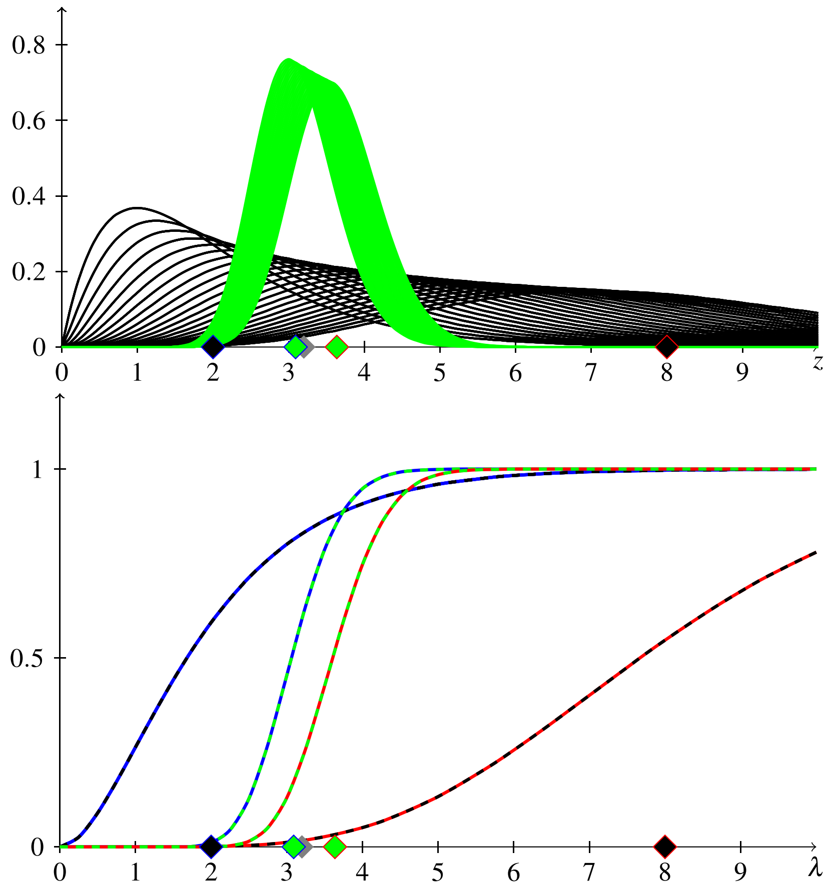

FIGURE 10. Lower and upper probability density function plot and pbox plot for the imprecise Gamma model with $s := 1$: prior (in black) and the corresponding posterior (in green).

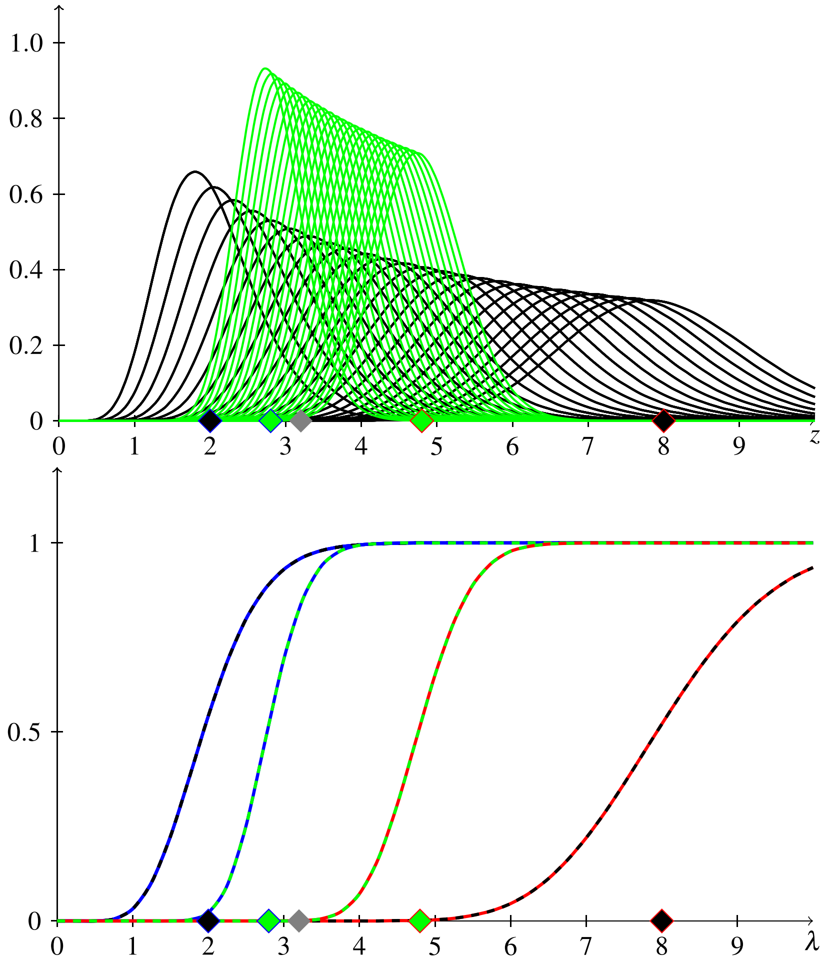

FIGURE 11. Lower and upper probability density function plot and pbox plot for the imprecise Gamma model with $s := 5$: prior (in black) and the corresponding posterior (in green).

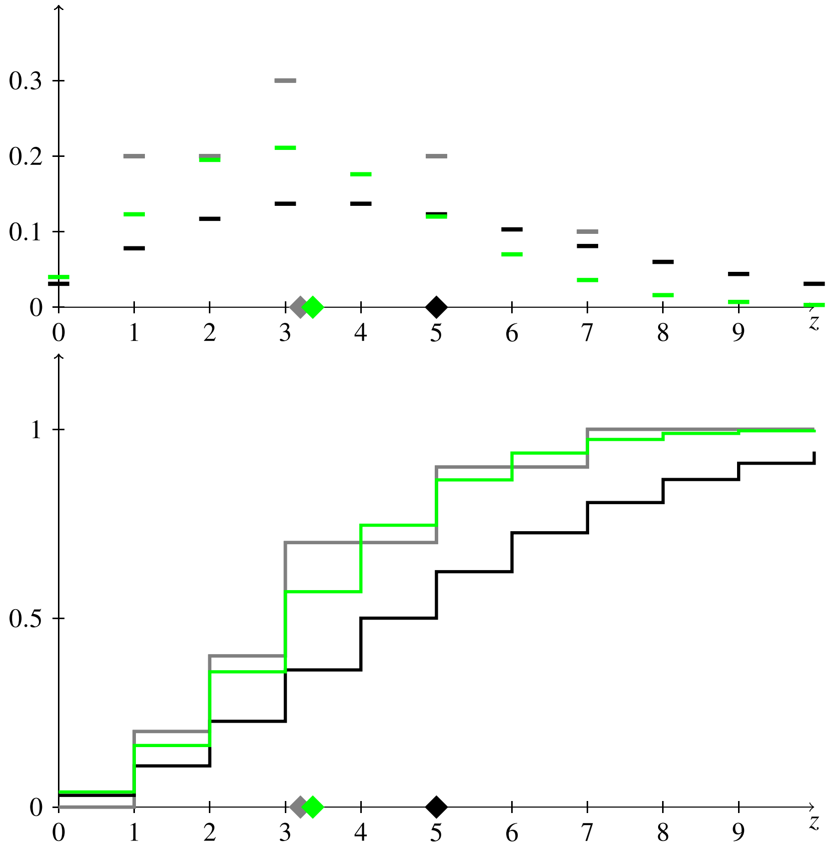

FIGURE 12. Probability mass function plot and cumulative distribution function plot for the Gamma-Poisson model with $s := 1$: prior (in black) and the corresponding posterior (in green).

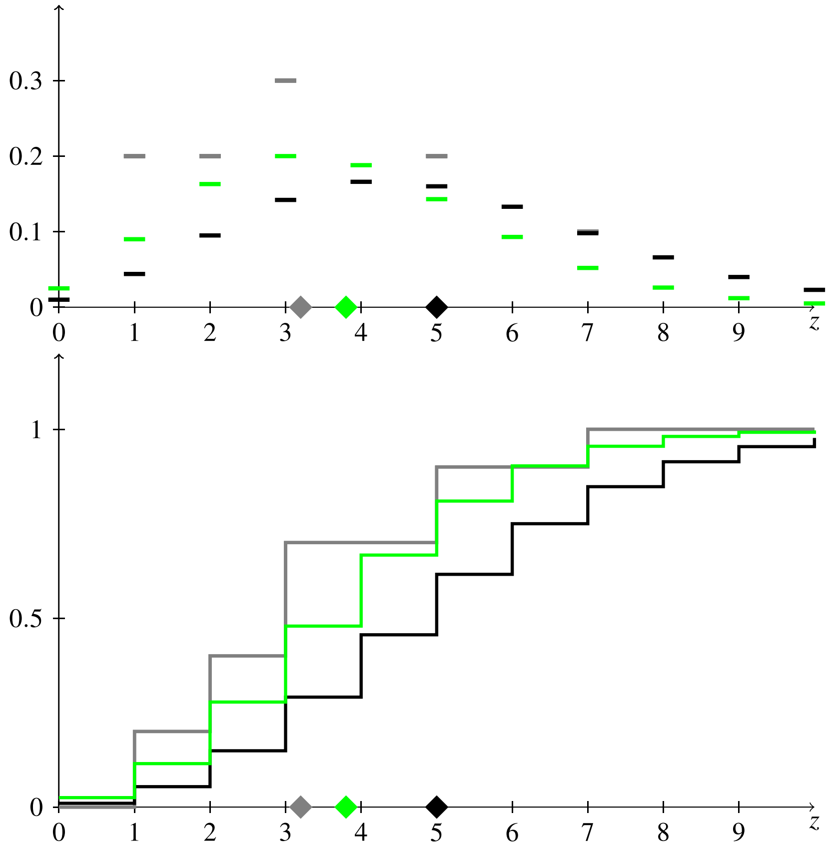

FIGURE 13. Probability mass function plot and cumulative distribution function plot for the Gamma-Poisson model with $s := 5$: prior (in black) and the corresponding posterior (in green).

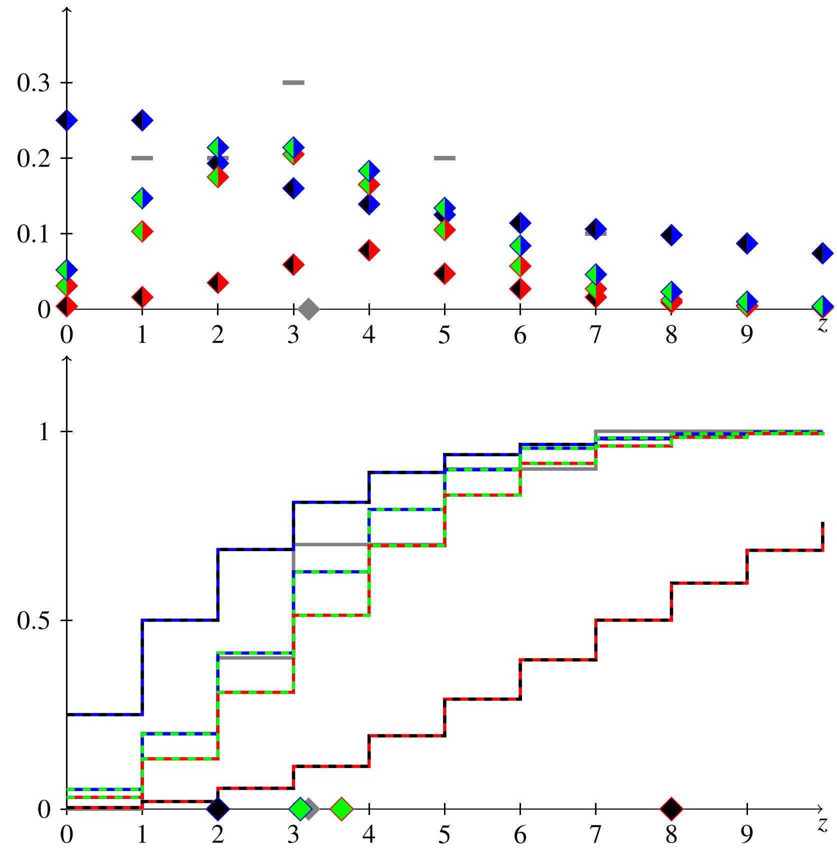

FIGURE 14. Lower and upper probability mass function plot and pbox plot for the imprecise Gamma-Poisson model with $s := 1$: prior (in black) and the corresponding posterior (in green).

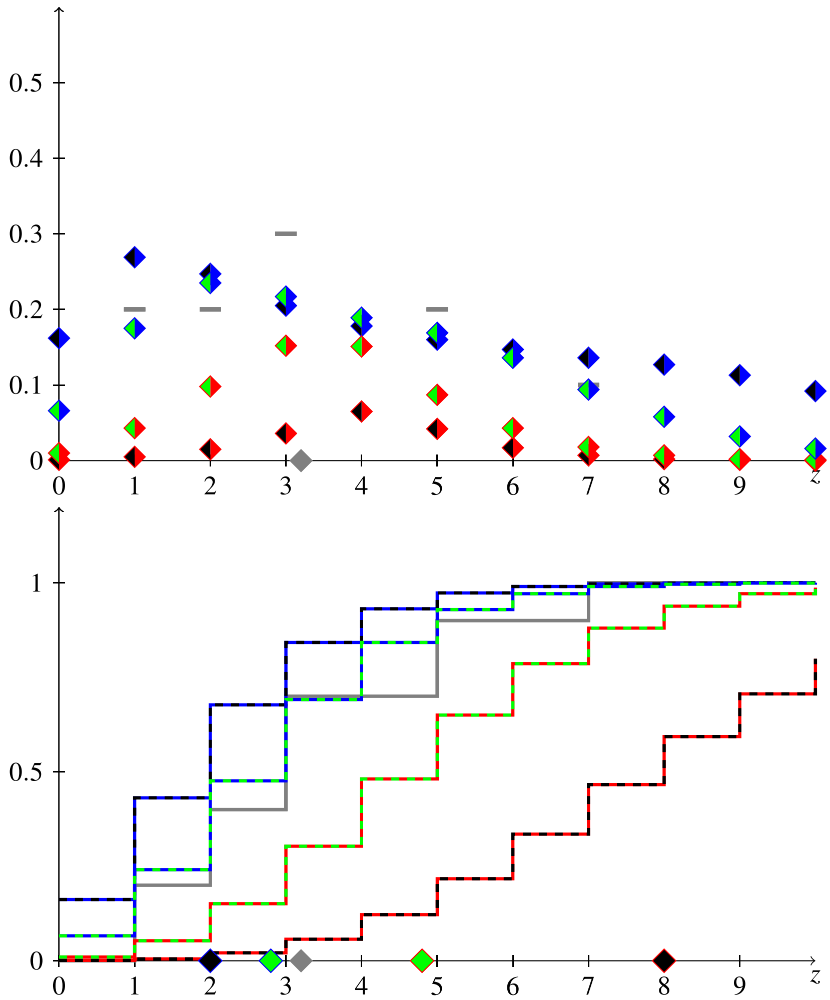

FIGURE 15. Lower and upper probability mass function plot and pbox plot for the imprecise Gamma-Poisson model with $s := 5$: prior (in black) and the corresponding posterior (in green).