Application Example 10
## (Joint, marginal and conditional distributions)

RELATION BETWEEN STORM DURATION AND PRECIPITATION INTENSITY

Hydrologists are much interested in the statistical characteristics of precipitation events. Such characteristics are for example used to assess the variability of river flow, evaluate the risk of floods, and design river regulation structures. The two most important storm characteristics are duration T and average precipitation intensity Y. The product TY gives the total precipitation volume V during the storm.

The random variables T and Y are probabilistically dependent. For example, the intensity Y is much more variable for short than for long-lasting storms. Whether a storm with characteristics (T, Y) is critical for a given river basin depends on the characteristics of the basin: large basins have a slow response and are mainly affected by long-lasting precipitation events, even though the intensity may have relatively low. By contrast, small basins have rapid response and for them, intense short storms produce the highest floods.

Simple models of precipitation in time use the joint distribution of T and Y to probabilistically describe the generic rainfall event as a "rectangular pulse" of random duration T and random intensity Y. The simplest models also assume that storms occur at Poisson times and ignore seasonality effects by assuming that (a) the rate λ of arrival of storms is constant throughout the year, and (b) the random vectors $(T,Y)_i$ that describe the characteristics of different storms i are independent and identically distributed. This is called the (stationary) Poisson Rectangular Pulse (or PRP) model; see Figure 1.
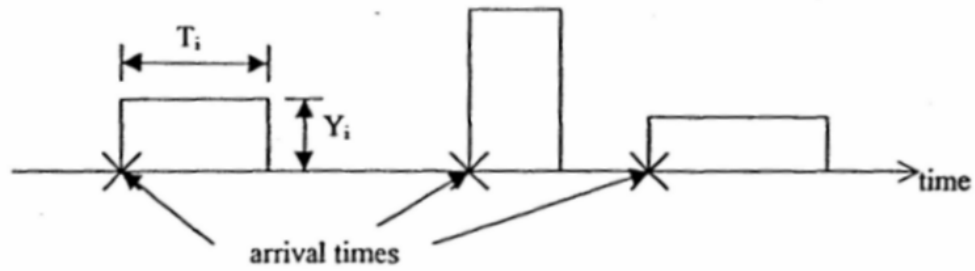
Figure 1: idealized rainfall model with "rectangular" rainfall events arriving at Poisson times

A large sample of storm durations and intensities has been collected at a raingage station in Bridgeport, Connecticut, where hourly rainfall data are available over a period of 11 years. A total of 1727 storms have been extracted from this record, as periods of continuous precipitation between hours when no precipitation occurs. A scatterplot of the observed (T,Y) values is shown in Figure 2. Notice that, as t increases, the dispersion of the conditional variable (Y|T=t) decreases.
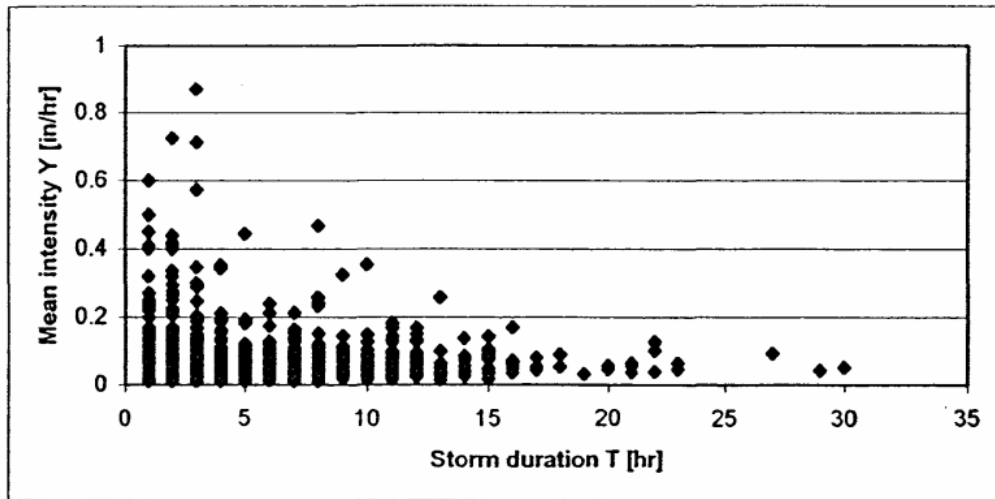
Figure 2: Scatterplot of storm duration T and mean precipitation intensity Y

In order to represent the rainfall process through a PRP model, we need the storm rate λ and the joint distribution of T and Y. The former can be directly estimated from the data as the number of observed storms divided by the period of observation. In our case this gives $\lambda = 1727/11 = 157$ storms/year.

The problem of estimating the joint distribution of T and Y is more difficult, especially because Y and T are dependent. Typically, joint distributions are obtained by estimating the marginal distribution of one variable, say T, and the conditional distributions of the other variable, here the conditional distribution of (Y|T=t) for each given t, and then using the relation

$$f_{T,Y}(t,y) = f_T(t)f_{Y|T=t}(y) \tag{1}$$

The marginal histogram of T from our rainstorm record is shown in Figure 3. It decays in a way that cannot be accurately represented by an exponential distribution. However, one may use a mixture of two exponential distributions, of the type

$$f_T(t) = 0.75\left(\frac{1}{2}e^{-t/2}\right) + 0.25\left(\frac{1}{7}e^{-t/7}\right) \tag{2}$$

This exponential mixture fits the empirical histogram very well; see Figure 3. One possible interpretation of Eq. 2 is that there are indeed two types of storms: 75% of the storms are generally short, with a mean duration of just 2 hours. These could be local convective storms. The remaining 25% of the storms have a mean duration of 7 hours and might be associated with the passage of frontal systems.
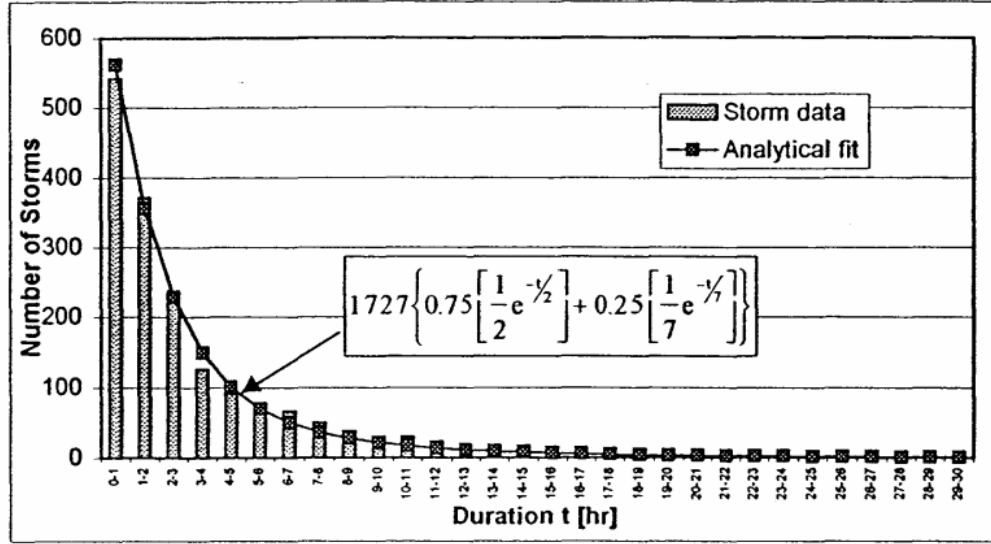
Figure 3: Empirical distribution of storm duration and fit by the exponential distribution mixture in Eq. 2

Estimation of the conditional distributions of (Y|T=t) for each given t is a more difficult task, but one can obtain a reasonable model with a modest amount of effort. The idea we exploit is that the distribution of (Y|T=t) might be of the same type for all t and simply have one or two parameters, like the mean value $m_{Y|t}$ and the standard deviation $\sigma_{Y|t}$, that depend on duration t. Empirical values of $m_{Y|t}$ and $\sigma_{Y|t}$ are shown in Figure 4 for each t. Notice that the mean precipitation intensity increases with increasing storm duration up to about 16 hours. For longer durations, there is some evidence that the mean intensity decreases. On the other hand, the standard deviation decreases monotonically with increasing storm duration. This trend of the standard deviation had already been observed on the scatterplot in Figure 2.

Except possibly for the intensity averages when T is large, the empirical values of $m_{Y|t}$ and $\sigma_{Y|t}$ are fitted well by the following functions (see Figure 4):

$$m_{Y|t} = 0.03 + 0.045(1 - e^{-t/3.5})$$
$$\sigma_{Y|t} = 0.08 e^{-t/18}$$

(3)

where t is in hours and Y is in inches/hour.
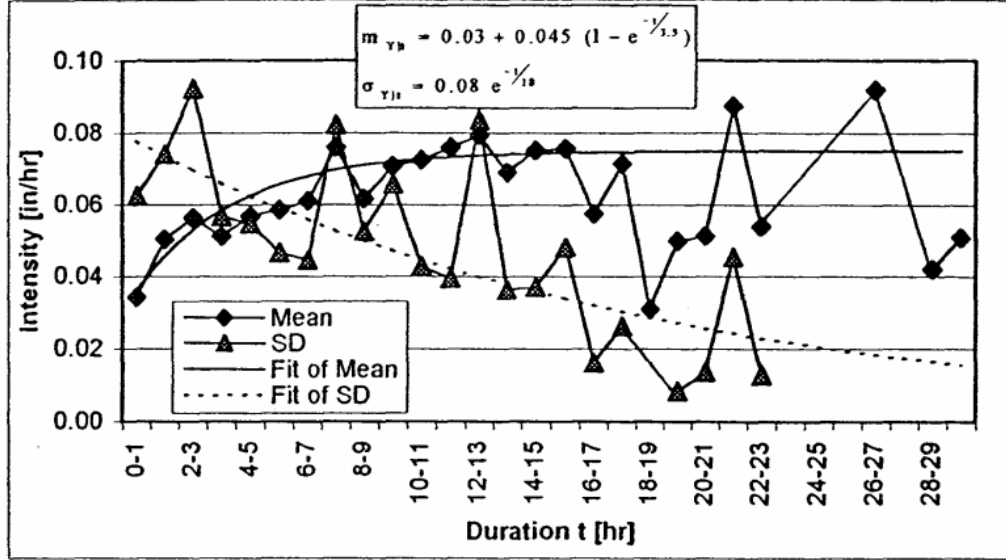
Figure 4: Empirical mean values and standard deviations of (Y|t) for different t and fitted analytic functions in Eq. 3

Finally, we need to specify the type of distribution of (Y|T). Figure 5 shows a representative subset of histograms of (Y|T) from the available sample. Visual inspection of the histograms in Figure 5 suggests that a lognormal distribution would provide a good fit. Recall that a random variable Y has lognormal distribution if its logarithm X = ln(Y) is normally distributed. The probability density function of a lognormal variable has the form

$$f_Y(y) = \frac{1}{y}\frac{1}{\sqrt{2\pi}\,\sigma_{\ln Y}}\, e^{-\frac{1}{2\sigma_{\ln Y}^2}\left[\ln(y)-m_{\ln Y}\right]^2}, \qquad y > 0 \tag{4}$$

where $m_{\ln Y}$ and $\sigma_{\ln Y}^2$ are the mean value and variance of ln(Y) and are related to the mean value and variance of Y as

$$\begin{aligned} m_{\ln Y} &= 2\ln(m_Y) - 0.5\ln(\sigma_Y^2 + m_Y^2) \\ \sigma_{\ln Y}^2 &= -2\ln(m_Y) + \ln(\sigma_Y^2 + m_Y^2) \end{aligned} \tag{5}$$

The relations in Eq. 5 may be used with the mean values and standard deviations of (Y|t) in Eq. 3 to obtain the parameters of the distribution of (Y|t) in Eq. 4, for any given t. Conditional probability density functions of (Y|T) obtained by this procedure are shown in Figure 5, superposed on the empirical histograms. The fit is quite good.

Finally, the joint density function $f_{T,Y}(t,y)$ is obtained as the product of the marginal density of T in Eq. 2 and the conditional density of Y in Eq. 4 (the latter, with parameters from Eqs. 3 and 5).
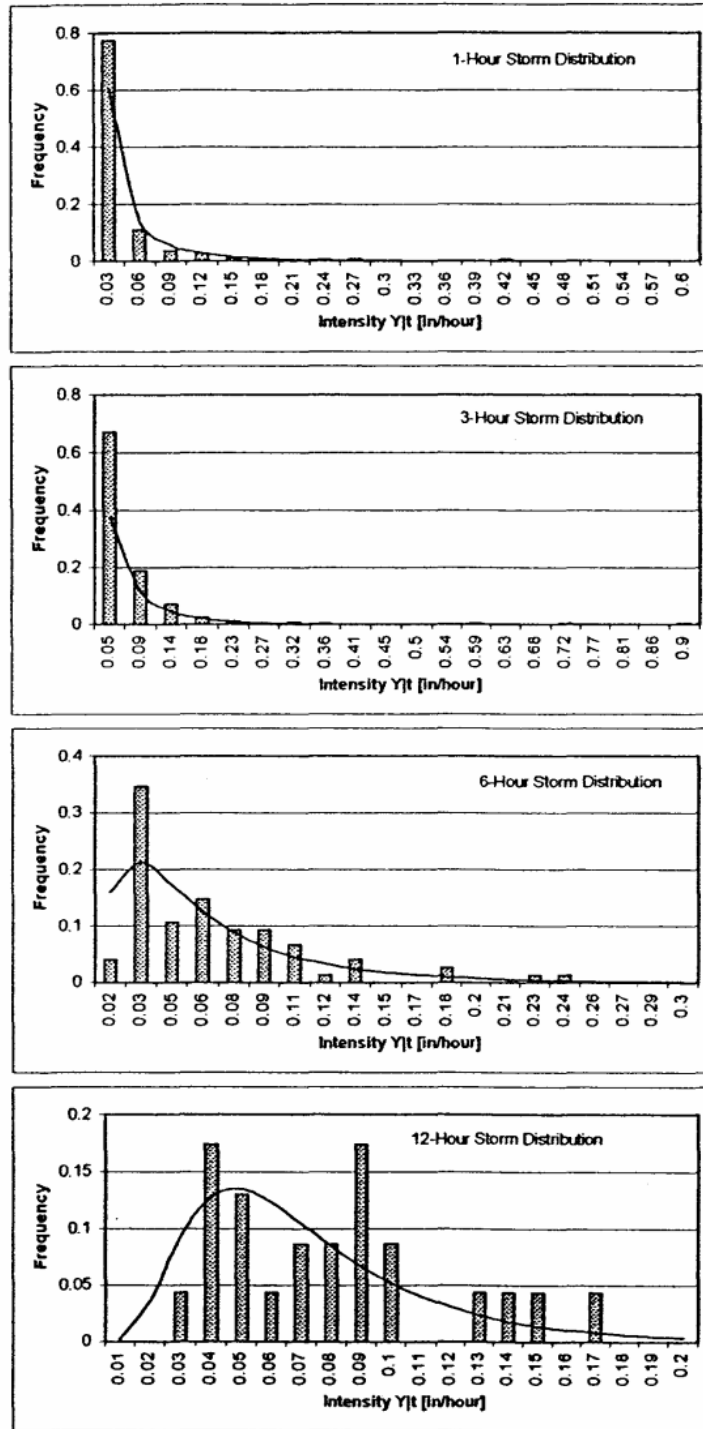
Figure 5: Selected histograms of the conditional variable (Y|t) and fitted lognormal distributions with parameters from Fig. 3

Problem 10.1

Using the Poisson Rectangular Pulse model and the joint distribution of *T* and *Y* described above, calculate numerically:

(a) the marginal distribution of *Y* for a generic storm

(b) the distribution of the total precipitation volume *Q = TY* for a generic storm;

(c) the probability that the maximum precipitation intensity in one year, Y_{max}, exceeds 1 inch/hour and the probability that the maximum precipitation volume in one year, Q_{max}, exceeds 10 inches.