1.010 Uncertainty in Engineering
Fall 2008

Application Example 11

(Functions of several random variables)

(Note that the Extreme Type Distribution will be covered in more detail in lectures relating to distribution models)

DISTRIBUTION OF THE MAXIMUM OF INDEPENDENT IDENTICALLY-DISTRIBUTED VARIABLES

Many engineering applications require the calculation of the distribution of the maximum of a number n of indendent, identically distributed (iid) variables. A typical situation is the design of a system for the “n-year demand” when the maximum demands in different years are iid (design of a dam for the n-year flood, design of an offshore platform for the n-year wave, design of a building for the n-year wind, etc.).

In some cases, for example the design of buildings against earthquake loads, using the year as the basic unit of time makes little sense, since earthquake occurrences do not have a yearly cycle (floods, winds, and sea-states do). Rather, earthquakes may be viewed as occurring at random times, say according to a Poisson process and maximization of the quantity of interest (for example earthquake magnitude or the induced monetary loss) should be done over the random number of earthquakes in a time period of duration T. Accordingly, we consider below maximum problems of two types:

$$Y_1 = \max\{X_1, X_2, ..., X_n\} \tag{1a}$$

$$Y_2 = \max\{X_1, X_2, ..., X_N\} \tag{1b}$$

where n in Eq. 1a is fixed (e.g. number of years) and N in Eq. 1b is a random variable with Poisson distribution and mean value λT.

Maximum of a fixed number n of iid variables (Eq. 1a)

Let $F_X(x)$ be the common distribution of the variables X_i in Eq. 1a and let $F_n(y)$ be the corresponding distribution of $Y_1 = \max\{X_1, X_2, ..., X_n\}$. Obtaining $F_n(y)$ from $F_X(x)$ is very simple. In fact,

$$F_n(y) = P[(X_1 \le y) \cap (X_2 \le y) \cap ... \cap (X_n \le y)] = \{F_X(y)\}^n \qquad (2)$$

Therefore, the CDF of Y_1 is obtained by taking the n^{th} power of the CDF of the X_i.

This result suffices when the distribution F_X is accurately known. In some cases, F_X is not completely known. It is then of interest to see whether, for large n, the distribution of Y_1 approaches a standard shape, which does not depend on F_X. Theoretical analysis shows that this indeed happens, but that the distribution $F_n(y)$ for n large is not entirely independent of F_X. One important result is that the distribution of Y_1 approaches a so-called Extreme Type 1 (EX1) distribution if the probability density of X decays in the upper tail as an exponential function. This includes exponential, normal, lognormal and gamma F_X distributions, among others. A second result is that, if the upper tail of X decays as a power function of x, then the distribution of Y_1 approaches a so-called Extreme Type 2 (EX2) distribution.

The EX1 and EX2 distributions have cumulative distribution functions of the type:

$$\text{EX1:} \qquad F(y) = e^{-e^{-\alpha(y-u)}}, \qquad -\infty < y < \infty, \quad \alpha > 0 \qquad (3a)$$

$$\text{EX2:} \qquad F(y) = e^{-(y/u)^{-k}}, \qquad y \ge 0, \quad u > 0, k > 0 \qquad (3b)$$

where α and u in Eq. 3a are parameters of EX1 and u and k in Eq. 3b are parameters of EX2. If a statistical sample of Y_1 or Y_2 is available, then these parameters may be estimated so that the theoretical mean value and variance of the distribution match the sample mean and sample variance. To use this moment-matching method of parameter estimation, one needs expressions for the mean and variance in terms of the parameters. These are:

For EX1: $m = u + \dfrac{0.577}{\alpha}$, $\sigma^2 = \dfrac{1.645}{\alpha^2}$ (4a)

For EX2: $m = u\Gamma\left(1-\dfrac{1}{k}\right), k > 1,$ $\sigma^2 = u^2\left[\Gamma\left(1-\dfrac{2}{k}\right) - \Gamma^2\left(1-\dfrac{1}{k}\right)\right], k > 2$ (4b)

where $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ is the so-called gamma function. Note that for EX2, the mean value diverges for $k \geq 1$ and the variance diverges for $k \geq 2$. While estimation of the parameters of EX1 is direct (first find α from the sample variance and then find u from the mean), direct estimation of the parameters of EX2 requires solving a system of nonlinear equations.

A simpler way to obtain the parameters of EX2 is to use the fact that, if a variable Y has EX2 distribution, then ln(Y) has EX1 distribution, with parameters $u_o = \ln(u)$ and $\alpha = k$. Therefore, one can take the natural log of the data, find their mean and variance, estimate u_o and α using Eq. 4a, and then obtain u and k from $u = e^{u_o}$ and $k = \alpha$.

Alternatively one may use the fact that the square of the coefficient of variation of EX2, $V^2 = \sigma^2/m^2$, depends only on the parameter k:

$$V^2 = \frac{\Gamma\left(1-\frac{2}{k}\right)}{\Gamma^2\left(1-\frac{2}{k}\right)} - 1 \qquad (5)$$

Therefore one may use Eq. 5 to find k and then the expression for m in Eq. 4b to find u.

The EX1 and EX2 distributions may be appropriate not just as models for the maximum values Y_1 and Y_2, but also for X. Going back to the examples of maximum floods, winds or sea-states, you may notice that such maximum values in year i, X_i, are themself the maxima of many random variables (for example, of 12 monthly maximum floods or sea-states). Therefore, the X_i themselves may be expected to have EX1 or EX2 distribution. The previous procedure to estimate the distribution parameters is most frequently applied to this case, because statistical samples are typically available for X (e.g. a record of yearly maximum floods, a record of yearly maximum winds) rather than Y (for which one would need a sequence of maximum n-year floods or winds).

Rate of Convergence to Extreme Distributions

One may wonder how fast the distribution of the maximum of n iid variables $\{X_1, ..., X_n\}$ converges to an extreme-type distribution. This depends on the distribution F_X of the variables X_i. To exemplify, suppose that the variables X_i are iid with exponential distribution and mean value 1; hence $F_X(x) = 1 - e^{-x}$. In this case the maximum is attracted to an EX1 distribution. From Eq. 2, the exact distribution is

$$F_n(y) = \left\{1 - e^{-y}\right\}^n \tag{6}$$

The mean value and standard deviation of this distribution depend on n and are displayed in Figure 1.

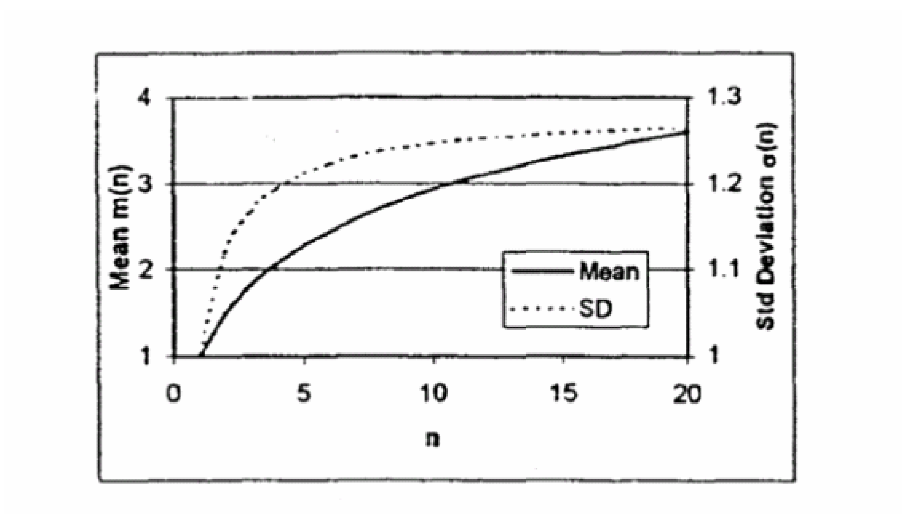

Figure 1: mean and standard deviation of $F_n(y)$ in Eq. 6 for different n

For selected values of n, Figure 2 compares the distribution in Eq. 6 with the EX1 distribution that has the same mean value and standard deviation. The parameters u and α of the EX1 distribution have been obtained using Eq. 4a and are shown in Figure 1.

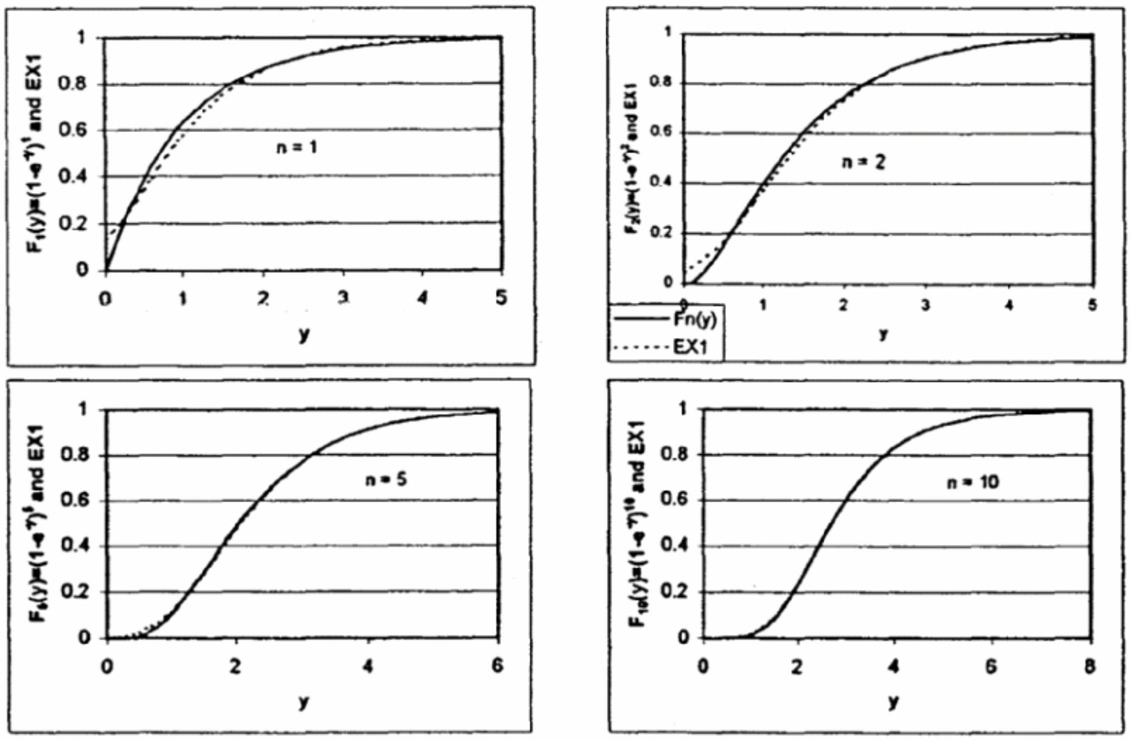

Figure 2: comparison for different n of the exact distribution $F_n(y)$ in Eq. 6 with the EX1 distribution having the same mean and variance

As Figure 2 shows, the exponential distribution has a shape that does not differ much from that of an EX1 distribution. Therefore, convergence to the EX1 distribution is quite rapid (for n = 10, the exact distribution is virtually identical to the approximating EX1 distribution).

As a second example, consider the case when the X_i have lognormal distribution. Specifically, suppose that the variables $Z_i = \ln(X_i)$ have standard normal distribution, with CDF $F_Z(z) = \Phi(z)$, where Φ is a symbol for the standard normal CDF. This means that the X_i have distribution $F_X(x) = \Phi(\ln x)$ and, from Eq. 2,

$$F_n(y) = \{\Phi(\ln y)\}^n \tag{7}$$

Again, we calculate the mean and standard deviation of Fn(y) as we have done for the exponential case. A plot of these quantities for different n is shown in Figure 3. Then we use these mean and variance values to fit an EX1 distribution. For selected values of n, Figure 4 compares the exact distribution in Eq. 7 with the approximating EX1 distribution.

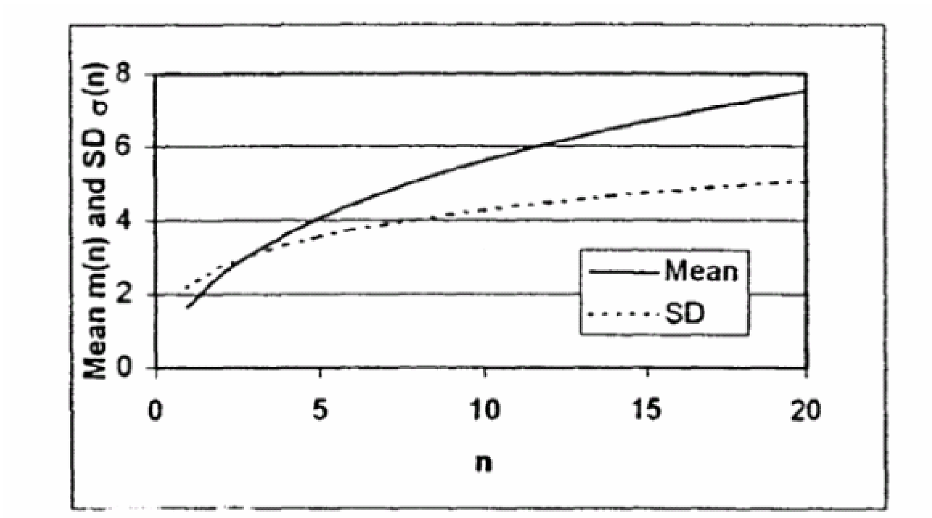

Figure 3: mean and standard deviation of $F_n(y)$ in Eq. 7 for different n

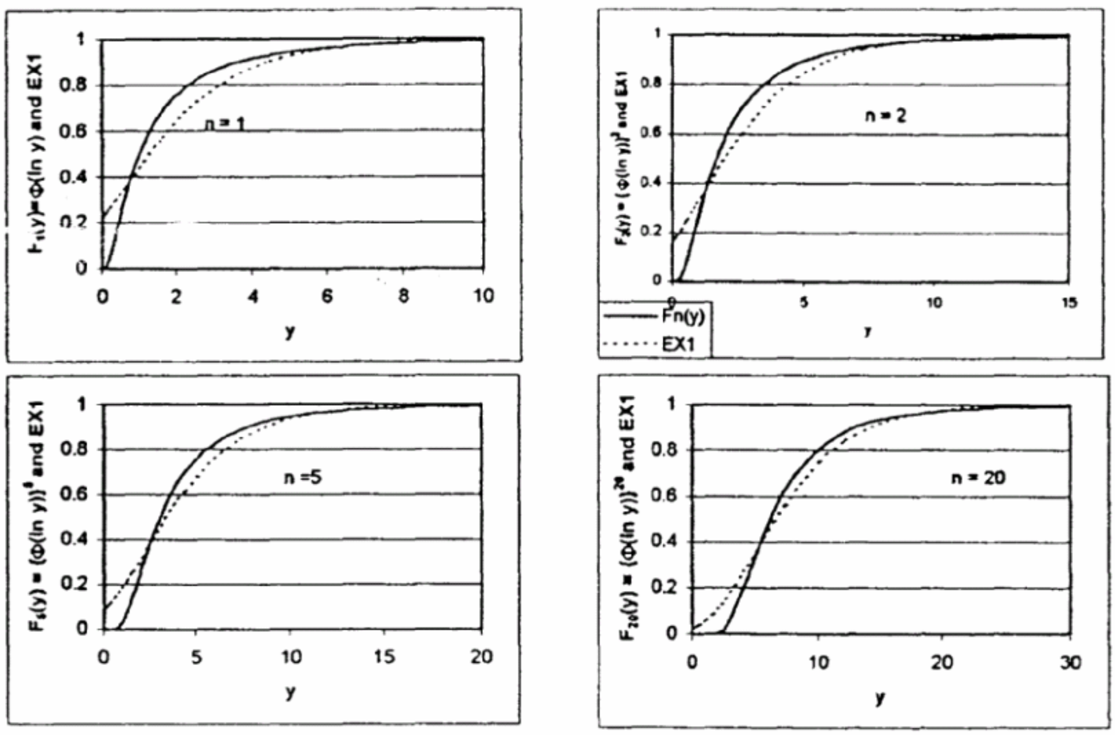

Figure 4: comparison of the exact distribution $F_n(y)$ in Eq. 7 with EX1 distributions having the same mean and variance

In this case, the convergence of $F_n(y)$ to an EX1 distribution is much slower and significant differences exist also for n = 20. The important message is that, when considering the maximum of iid variables, convergence is not as rapid as that of sums of iid variables to the normal distribution. Therefore, one should be cautious in assuming an extreme value distribution when n is not large. Whenever possible, one should use the exact expression for $F_n(y)$ in Eq. 2.

Problem 11.1

Make an analysis similar to the previous two examples for the case when X has standard normal distribution.

The Maximum of a Poisson number N of iid variables (Eq. 1b)

Consider now the maximum of a Poisson number N of variables (Eq. 1b). It is again easy to obtain exact results for any given distribution F_X of the X variables and any given mean value λT of N. This is done as follows.

Let $F_N(y)$ be the distribution of Y_2 in Eq. 1b. Then $F_N(y)$ is the probability that none of the N events in T has intensity X greater than y. Events with this characteristic occur according to a Poisson process with reduced rate $\lambda_y = \lambda[1 - F_X(y)]$. Therefore, the probability that no such event occurs in T is $e^{-\lambda_y T} = e^{-\lambda T[1-F_X(y)]}$ and

$$F_N(y) = e^{-\lambda T[1-F_X(y)]} \tag{8}$$

For example, if X has exponential distribution $F_X(x) = 1 - e^{-x/m}$,

$$F_N(y) = e^{-\lambda T e^{-y/m}}, \qquad y \geq 0 \tag{9}$$

Notice that, for y = 0, Eq. 9 gives $F_N(0) = \exp\{-\lambda T\}$, which is the probability of no event in T. For y > 0, Eq. 9 has the form of the EX1 distribution in Eq. 3a, with $\alpha = 1/m$ and $u = m\ln(\lambda T)$. We conclude that, in the present case of exponentially distributed X variables, Y_2 in Eq. 1b has a distribution of mixed type (neither entirely discrete nor entirely continuous). The distribution has a probability mass $\exp\{-\lambda T\}$ at the origin and the rest of the distribution has the form of a “truncated EX1 distribution”.

Problem 11.2

For the seismic design of structures, one needs to find the distribution of the maximum earthquake magnitude in T years in a given region. Potentially damaging earthquakes (say, earthquakes of magnitude greater than 5) occur with good approximation according to a Poisson process with rate $\lambda_{>5}$ *and have independent and exponentially distributed magnitudes,*

$$F_M(m) = 1 - e^{-(m-5)/m_o}, \qquad m \geq 5, \; m_o = 0.3 \qquad (10)$$

Notice that this is a shifted exponential distribution with 5 as minimum possible value and that m is used as a symbol for magnitude, not for mean value.

(a) Using results given above, find the distribution of the maximum magnitude in T years, as a function of T and the Poisson rate $\lambda_{>5}$*;*

(b) The distribution you obtained in Part (a) should depend on T and $\lambda_{>5}$ *only through the product* $\lambda_{>5}T$*, which is the expected number of earthquakes of magnitude above 5 in T years. Plot the cumulative distribution of the maximum magnitude in T years for* $\lambda_{>5}T$ *= 0.1, 1, 10. Comment on the results.*