SPECIAL CONTINUOUS DISTRIBUTIONS

Uniform

X can have values only in an (a, b) interval, all those values are 'equally likely'. This means: $f(x) = \dfrac{1}{b-a}$ when $a < x < b$. $F(x) = \dfrac{x-a}{b-a}$ (same range) $\mathbb{E}(X) = \dfrac{1}{b-a}\int\limits_a^b x\,dx = \dfrac{a+b}{2}$ (we could have used symmetry), $\mathbb{E}(X^2) = \dfrac{1}{b-a}\int\limits_a^b x^2\,dx = \dfrac{a^2+ab+b^2}{3}$, implying: $\mathrm{Var}(X) = \dfrac{a^2+ab+b^2}{3} - \left(\dfrac{a+b}{2}\right)^2 = \dfrac{(b-a)^2}{12}$ and

$$\sigma = \frac{b-a}{2\sqrt{3}}$$

The $\mu \pm \sigma$ interval of length $\dfrac{b-a}{\sqrt{3}}$ contains $\dfrac{1}{\sqrt{3}} = 57.74\%$ of total probability ($\mu \pm 2\sigma$ already covers the whole range). Symbolic notation: $\mathcal{U}(a, b)$.

Exponential

distribution relates to the 'arriving customers' experiment, where λ (say, per minute) is the average arrival rate. This time, X is the time from now (set to time 0) till the arrival of the first customer. To find its PDF, we subdividing each *minute* into k subintervals, assuming that the probability of an arrival during any one of these is $p = \frac{\lambda}{k}$. We then introduce a *geometric*-type RV Y which *counts* the *number* of these *subintervals* till the first arrival. Obviously

$$\Pr(X \le x) \simeq \Pr(Y \le xk) = 1 - \left(1 - \tfrac{\lambda}{k}\right)^{xk}$$

where $k = 0, 1, 2, \ldots$ All we need to do to get the distribution function of X is to take the limit of the last expression as $k \to \infty$. Thus

$$F(x) = 1 - e^{-x\lambda}$$

when $x > 0$. In this context, it is common to introduce a new parameter $\beta = \frac{1}{\lambda}$ (the average time between consecutive arrivals), which means that

$$F(x) = 1 - \exp(-\tfrac{x}{\beta})$$

Based on this

$$f(x) = F'(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$$

when $x > 0$. We can now compute $M(t) = \frac{1}{\beta}\int\limits_0^\infty e^{-\frac{x}{\beta}} \cdot e^{tx}\, dx = \frac{1}{\beta}\int\limits_0^\infty e^{-x(\frac{1}{\beta}-t)}\, dx = \frac{1}{\beta(\frac{1}{\beta}-t)} = \frac{1}{1-\beta\, t}$, whose Taylor expansion (in terms of t) is $1 + \beta t + \beta^2 t^2 + \beta^3 t^3 + \beta^4 t^4 + ...$ This yields $\mu = \beta$, $\mathrm{Var}(X) = 2\beta^2 - \beta^2 = \beta^2$, $\sigma = \beta$, skewness: $\frac{\mathbb{E}(X^3)-3\mu\mathbb{E}(X^2)+2\mu^3}{\sigma^3} = \frac{6\beta^3-3\times 2\beta^3+2\beta^3}{\beta^3} = 2$, and kurtosis $\frac{\mathbb{E}(X^4)-4\mu\mathbb{E}(X^3)+6\mu^2\mathbb{E}(X^2)-3\mu^4}{\sigma^4} = \frac{24\beta^4-4\times 6\beta^4+6\times 2\beta^4-3\beta^4}{\beta^4} = 9$.

The exponential distribution shares, with the geometric distribution (from which it was derived), the '`memory-less`' property of $\Pr(X - a > x | X > a) = \Pr(X > x)$, i.e. given we have been waiting, *unsuccessfully*, for time a, the probability that the first arrival will take longer than x *from now* is the same as when we started waiting. `Proof`: $\Pr(X - a > x | X > a) = \frac{\Pr(X>x+a \,\cap\, X>a)}{\Pr(X>a)} = \frac{\Pr(X>x+a)}{\Pr(X>a)} = \frac{\exp\left(-\frac{x+a}{\beta}\right)}{\exp\left(-\frac{a}{\beta}\right)} = \exp\left(-\frac{x}{\beta}\right) = \Pr(X > x)$. ■

The potential applications of this distribution include: time intervals between consecutive phone calls, accidents, fishes caught, etc. (all those discussed in connection with the Poisson distribution).

$\mathcal{E}(\beta)$ will be our symbolic notation for the exponential distribution with the mean of β.

Another important characteristic of a continuous distribution is the `median`, denoted $\tilde{\mu}$ and defined as a solution to $F(\tilde{\mu}) = \frac{1}{2}$ (the point which 'splits' the PDF in two halves, each having the same 50% probability). For the exponential distribution, it equals to $\beta \ln 2 = 0.6931\beta$, which is substantially smaller than μ. This means that: if it takes, on the average, 1 hour to catch a fish, 50% of all fishes are caught in less than 41 min. and 35 sec.(not a contradiction).

Consider n *independent* RVs (say , X_1, X_2,....X_n), all exponentially distributed, with the same mean β. What is the distribution of $Y \equiv \min(X_1, X_2, ...X_n)$? $\triangleright$ $\Pr(Y > y) = \Pr(X_1 > y \cap X_2 > y \cap .... \cap X_n > y) = \Pr(X_1 > y) \cdot \Pr(X_2 > y) \cdot ... \cdot \Pr(X_n > y) = \left(e^{-\frac{y}{\beta}}\right)^n = e^{-\frac{n\,y}{\beta}}$. This implies that the distribution function of Y equals $1 - e^{-\frac{n\,y}{\beta}}$ when $y > 0$. This clearly identifies the distribution of Y as *exponential* with the mean of $\frac{\beta}{n}$ (things happen n times faster now). When the individual means are distinct (say: β_1, β_2, ...β_n) the result is (similarly) an exponential distribution with the mean of

$$\frac{1}{1/\beta_1 + 1/\beta_2 + ... + 1/\beta_n}$$

Gamma

distribution relates to the previous experiment, except now X is the time of the k^{th} arrival (from 'now', when we set time to 0). X is clearly a *sum* of k independent RVs of the *exponential* type, each having the mean of β.This immediately implies that $\mathbb{E}(X) = k\beta$, $\mathrm{Var}(X) = k\beta^2$ and $M(t) = \frac{1}{(1-\beta\, t)^k}$. Going back to our discussion of the *Poisson* distribution, we realize that we already know

$$\Pr(X > x) = \left(1 + \lambda x + \frac{(\lambda x)^2}{2} + \frac{(\lambda x)^3}{3!} + ... + \frac{(\lambda x)^{k-1}}{(k-1)!}\right) e^{-\lambda x}$$

(the probability of fewer than k arrivals before time x), where $\lambda = \frac{1}{\beta}$. Differentiating $F(x) = 1 - \Pr(X > x)$ with respect to x yields:

$$\frac{x^{k-1} e^{-\frac{x}{\beta}}}{\beta^k (k-1)!}$$

when $x > 0$. We can verify the correctness of this answer by computing the corresponding MGF: $M(t) = \frac{1}{\beta^k (k-1)!} \int\limits_0^\infty x^{k-1} e^{-\frac{x}{\beta}} \cdot e^{xt}\, dx = \frac{1}{\beta^k (k-1)!} \int\limits_0^\infty x^{k-1} e^{-x(\frac{1}{\beta}-t)}\, dx = \frac{1}{\beta^k (k-1)!} \cdot \frac{(k-1)!}{(\frac{1}{\beta}-t)^k} = \frac{1}{(1-\beta t)^k}$ (check).

EXAMPLES:

- If $X \in \gamma(4, 20\text{ min.})$, find $\Pr(X < 30\text{ min.}) = 1 - e^{-\frac{30}{20}}[1 + \frac{30}{20} + \frac{(\frac{30}{20})^2}{2} + \frac{(\frac{30}{20})^3}{3!}] = 6.56\%$, and $\Pr(X > 2\text{ hr.}) = e^{-\frac{120}{20}}[1 + 6 + \frac{6^2}{2} + \frac{6^3}{6}] = 15.12\%$. Or, by Maple, more directly

$$\begin{aligned}
\int_0^{30} \frac{x^3 \exp(-\frac{x}{20})}{3! \cdot 20^4} &= 6.564\% \\
\int_{120}^{\infty} \frac{x^3 \exp(-\frac{x}{20})}{3! \cdot 20^4} &= 15.12\%
\end{aligned}$$

- A fisherman whose average time for catching a fish is 35 minutes wants to bring home exactly 3 fishes. What is the probability he will need between 1 and 2 hours? $\triangleright$ $\Pr(1\text{ hr.} < X < 2\text{ hrs}) = F(120\text{ min.}) - F(60\text{ min.}) = e^{-\frac{60}{35}}[1 + \frac{60}{35} + \frac{(\frac{60}{35})^2}{2}] - e^{-\frac{120}{35}}[1 + \frac{120}{35} + \frac{(\frac{120}{35})^2}{2}] = 41.92\%$. More directly:

$$\int_{60}^{120} \frac{x^2 \exp(-\frac{x}{35})}{2 \cdot 35^3} = 41.92\%$$

- If a group of 10 fishermen goes fishing, what is the probability that the *second* catch of the *group* will take less than 5 min. Assume the value of $\beta = 20$ min. for *each* fisherman; also assume that the one who catches the first fish *continues* fishing. $\triangleright$ This is equivalent to having a single super-fisherman who catches fish at a 10 times faster rate, i.e. $\beta_{\text{group}} = 2$ min. Answer: $1 - e^{-\frac{5}{2}}(1 + \frac{5}{2}) = 71.27\%$. More directly:

$$\int_0^5 \frac{x \cdot \exp(-\frac{x}{2})}{2^2} = 71.27\%$$

For large k, the gamma distribution is approximately Normal.

EXAMPLE: Phone calls arrive at the rate of 12.3/hour. What is the probability that the 50^{th} phone call will arrive after 1 p.m. if the office opens at 8 a.m. $\triangleright$ If X is the time of the arrival of the 50^{th} phone call, in hours (setting *our* time to 0 at 8 a.m.) we have to find $\Pr(X > 5\text{ hr.})$. The distribution of X is $\gamma(50, \frac{1}{12.3}\text{ hr.})$ which implies that the answer is approximately equal to $\frac{1}{\sqrt{2\pi \times 50/12.3}} \int_5^{\infty} \exp[-\frac{(x-50/12.3)^2}{2 \times 50/12.3^2}]dx = 5.19\%$. Alternate solution:

We can also introduce Y as the number of phone calls received during the 8 a.m.-1 p.m. time interval. Its distribution is Poisson, with $\Lambda = 5 \times 12.3 = 61.5$. The question can be reformulated as $\Pr(Y < 50) \simeq \frac{1}{\sqrt{61.5\times 2\pi}} \int_{-\infty}^{49.5} \exp[-\frac{(y-61.5)^2}{2\times 61.5}]dy = 6.30\%$. (the exact answer is 5.91%).

Central Limit Theorem

Let us investigate the distribution of

$$Z \equiv \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

which has the expected value of 0 and standard deviation of 1, for *any* value of n (so that we can take the limit $n \to \infty$).

Realizing that Z can be also written as $\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma \sqrt{n}} \right)$, we can easily construct its MGF (by finding the MGF of a single $\frac{X_i - \mu}{\sigma \sqrt{n}} \equiv Y$, and raising it to the power of n).

We know that, in general, $M_y(t) = 1 + \mathbb{E}(Y)t + \mathbb{E}(Y^2)\frac{t^2}{2} + \mathbb{E}(Y^3)\frac{t^3}{3!} + \mathbb{E}(Y^5)\frac{t^4}{4!} + ...$, which, in this particular case reads:

$$1 + \frac{t^2}{2n} + \frac{\alpha_3 t^3}{6n^{3/2}} + \frac{\alpha_4 t^4}{24n^2} + \ ...$$

where α_3 and α_4 is the skewness and kurtosis of the sampled distribution. This implies that

$$M_z(t) = \left(1 + \frac{t^2}{2n} + \frac{\alpha_3 t^3}{6n^{3/2}} + \frac{\alpha_4 t^4}{24n^2} + \ ... \right)^n$$

Taking the $n \to \infty$ limit, we get

$$M_z(t) \underset{n\to\infty}{\longrightarrow} e^{t^2/2}$$

Thus, we get a rather unexpected result: the distribution of Z has (for large enough n) the same symmetric shape, not in the least affected by the shape of the original distribution form which the sample is taken.

We now have to start looking for a RV whose MGF is $\exp(t^2/2)$.