



JORGE A. RAMÍREZ

Professor

Water Resources, Hydrologic and Environmental Sciences

Civil Wngineering Department

Fort Collins, CO 80523-1372

Phone: (970) 491-7621

FAX: (970) 491-7727

e-mail: Jorge.Ramirez@ColoState.edu

DERIVED DISTRIBUTION APPROACH

In the disciplines of science and engineering, relationships that predict the value of a dependent variable in terms of one or many basic (independent) variables are commonly developed. Physical systems are naturally complex. The functional form of the relationship between independent and dependent quantities, or the value of the independent variables (or both) are not known with certainty, always. Techniques based on probabilistic assumptions can be used to account for this uncertainty. When the uncertainty derives from uncertainty in the independent variables, but not from uncertainty in the functional dependence, a *derived distribution approach* leads to the probability density function (PDF) of the dependent variable. In this case, the functional form relating independent and dependent variables is assumed known with certainty. In such instances, it is possible to derive the PDF of the dependent variable from that [or those] of the independent variable [or variables] (Ang and Tang, 1975). The methodology employed to do so is commonly known as the *derived distribution approach*. There are numerous examples of applications of the derived distribution approach in hydrologic science and water-resources engineering (*e.g.*, Eagleson, 1972; Hebson and Wood, 1982; Diaz-Granados et al., 1987; Entekhabi and Eagleson, 1989; Ramirez and Senarath, 2000).

The underlying basic conceptual framework of the *derived distribution approach* can be very clearly exemplified by the following simple derivation for the case of a function of a single variable. This can be extended to multivariate cases, but the procedure can become very complex rapidly.

If a certain dependent variable, y is a function of a single, continuous, and independent variable, x then:

$$y = g(x) \quad (1)$$

Assume that $y = g(x)$ is a monotonically increasing function of x with a unique inverse equal to:

$$x = g^{-1}(y) \quad (2)$$

If the independent variable x is a realization of a random variable X , then the Cumulative Distribution Function (CDF) of Y can be obtained from the known CDF of X as follows:

$$F_Y(y) = P[Y \leq y] = P[X \leq g^{-1}(y)] = F_X[g^{-1}(y)] \quad (3)$$

However, since x is continuous, the CDF of Y can be written using the following integral form:

$$F_Y(y) = \int_{x \leq g^{-1}(y)} f_X(x) dx = \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \quad (4)$$

After changing the variable of integration, (4) can be re-written as follows:

$$F_Y(y) = \int_{-\infty}^y f_X[g^{-1}(y)] \frac{dg^{-1}(y)}{dy} dy \quad (5)$$

To obtain the PDF of Y , (5) needs to be differentiated with respect to y . The resulting relationship is as follows:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = f_X[g^{-1}(y)] \frac{dg^{-1}(y)}{dy} \quad (6)$$

Similarly, if $g(x)$ is a monotonically decreasing function of y with a unique inverse equal to (2), then the CDF of Y is:

$$F_Y(y) = 1 - F_X[g^{-1}(y)] \quad (7)$$

Moreover, the PDF of Y is obtained by differentiating (7) with respect to y :

$$f_Y(y) = -f_X[g^{-1}(y)] \frac{dg^{-1}(y)}{dy} \quad (8)$$

Therefore, in general the derived PDF of Y can be written as follows:

$$f_Y(y) = f_X[g^{-1}(y)] \left| \frac{dg^{-1}(y)}{dy} \right| \quad (9)$$

For clarity, substituting x for $g^{-1}(y)$, (9) can be re-written in the following form:

$$f_Y(y) dy = f_X(x) dx \quad (10)$$

Figure 1 provides a graphical interpretation of (10). As pointed out by Benjamin and Cornell (1970) with regard to (10), "the likelihood that Y takes on a value in an interval of width dy centered on the value y is equal to the likelihood that X takes on a value in an interval centered on the corresponding value $x = g^{-1}(y)$, but of width $dx = dg^{-1}(y)$ ".

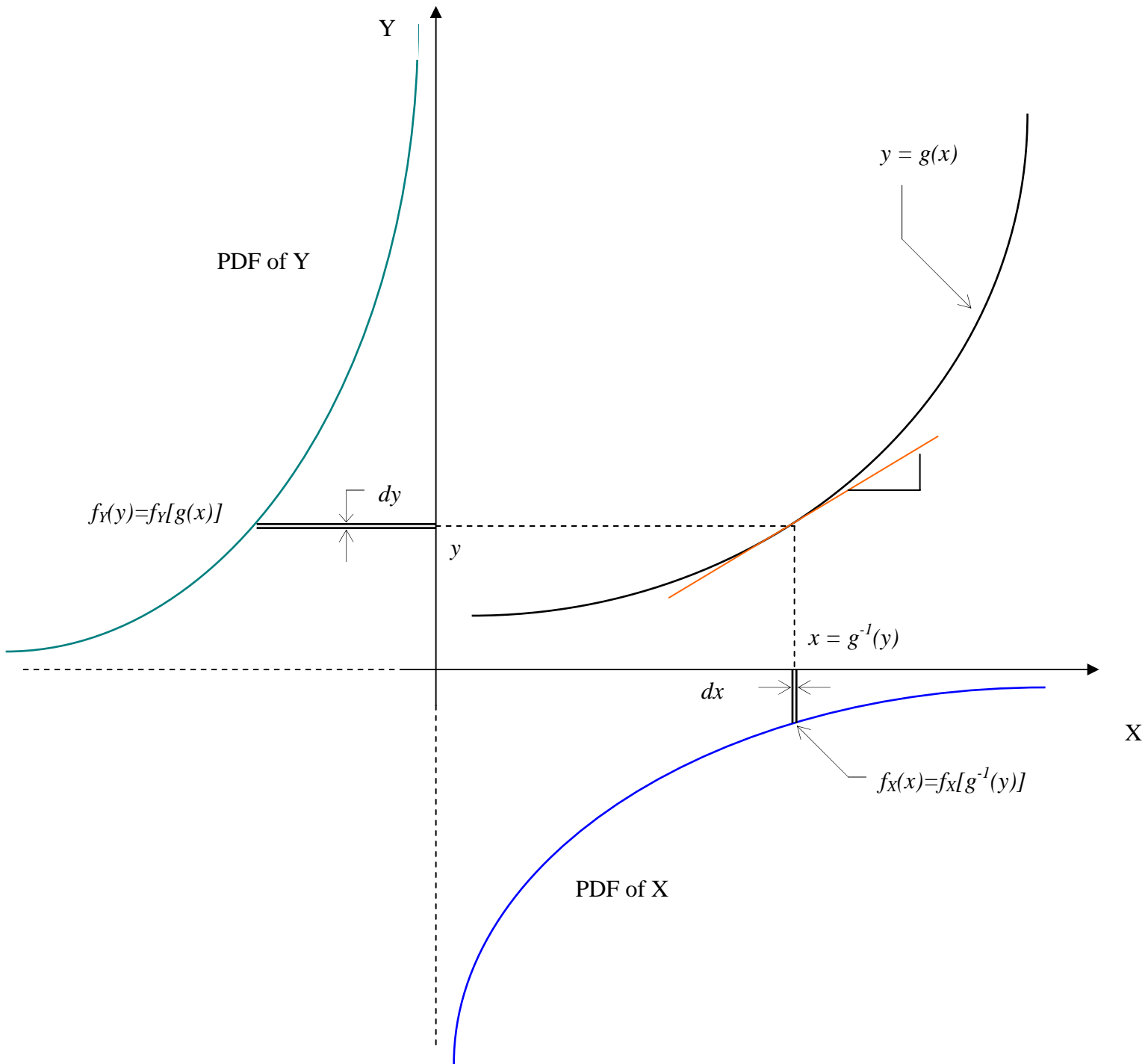


Figure 1: Schematic representation of the derived distribution approach.

For example. let $y(x)$ be:

$$y = ae^{-cx}, x \geq 0 \quad (11)$$

such that the pdf of X is:

$$f_X(x) = \alpha e^{-\alpha x}, \quad x \geq 0$$

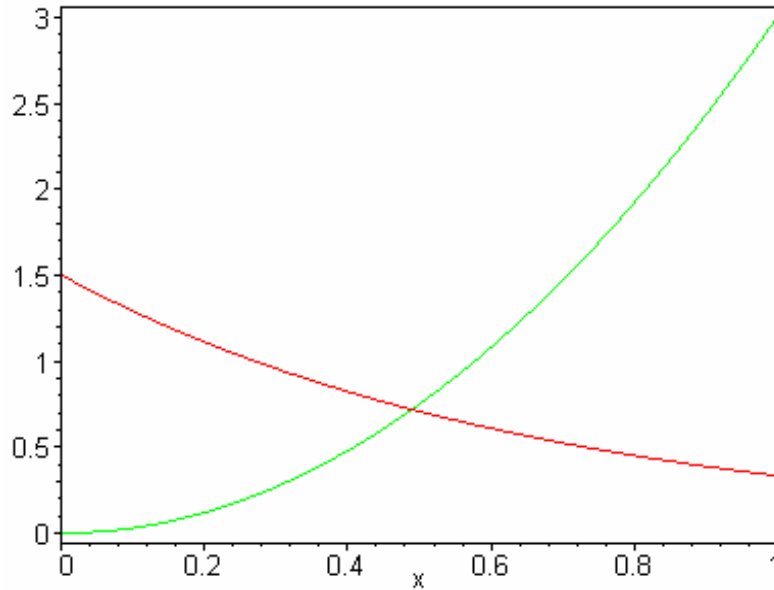
Using Equations 3 and 4, the cumulative distribution function (cdf) of Y can be expressed as,

$$F_Y(y) = \int_{-(1/c)\ln(y/a)}^{\infty} \alpha e^{-\alpha x} dx = (y/a)^{\alpha/c}, \quad 0 \leq y \leq a \quad (12)$$

Then, the pdf of Y can be derived as,

$$f_Y(y) = \frac{\alpha}{ac} \left(\frac{y}{a}\right)^{\frac{\alpha}{c}-1}, \quad 0 \leq y \leq a; \quad \alpha, a, c \geq 0 \quad (13)$$

whose mean is $E[Y] = a \alpha / (\alpha + c)$.



Derivation of the log-normal probability density function of Q

Let Y be a Gaussian random variable with distribution $N(\mu_Y, \sigma_Y^2)$. Therefore, the probability distribution of $Y, f_Y(\cdot)$ is

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \quad (14)$$

Now, let Q be a random variable such that $Q=e^Y$ (that is, $Y=\ln(Q)$). Therefore, the probability distribution of $Q, f_Q(\cdot)$, is a log-normal distribution (by definition, the probability distribution of $Q, f_Q(\cdot)$, is log-normal if the distribution of $Y=\ln(Q)$ is normal.)

Using the notation from the attached document on derived distributions, we see that $Q = e^Y = g(Y)$ and $Y = \ln(Q) = g^{-1}(Q)$. Therefore,

$$\frac{dg^{-1}(Q)}{dQ} = \frac{1}{Q}$$

Applying the derived distribution approach to obtain the distribution of Q from knowledge of the distribution of Y , we use equation 9

$$f_Q(Q) = f_Y[g^{-1}(Q)] \frac{dg^{-1}(Q)}{dQ}$$

to obtain the log-normal probability density function of Q as,

$$f_Q(Q) = \frac{1}{Q\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(\ln Q - \mu_Y)^2}{2\sigma_Y^2}} \quad (15)$$

In the literature, the following expression is often given as the pdf of Q when Q is log-normally distributed,

$$f_Q^*(Q) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(\ln Q - \mu_Y)^2}{2\sigma_Y^2}} \quad (16)$$

However, it can easily be shown that the function of (16) is not a proper probability density function because $\int_{-\infty}^{\infty} f_Q^*(Q) dQ \neq 1$. For example, for $\mu_Y=0.5$ and $\sigma_Y=1$ that integral is equal

to 2.718. However, the distributions of (14) and (15) are proper probability density functions,

that is, $\int_{-\infty}^{\infty} f_Y(y)dy = 1$ and $\int_{-\infty}^{\infty} f_Q(Q)dQ = 1$.

Finally, the figures below show how these three functions compare to each other. The graphs shown correspond to $\mu_Y=0.5$ and $\sigma_Y=1$.

