Important Distributions/Functions of Random Variables

ECE 313

Probability with Engineering Applications

Lecture 23 - November 15, 1999

Professor Ravi K. Iyer

University of Illinois

Functions of a Random Variable

- Knowledge of some characteristic of the system, with knowledge of the input, allows some estimate of the behavior at the output.
- E. g., the input random variable *X* and its density *f(x)* are known and the input-output behavior is characterized by:

$$Y = \Phi(X)$$

- To compute the density of the random variable *Y.*
- We assume that Φ is continuous or piecewise continuous, so $Y = \Phi(X)$ will be a random variable.

Functions of a Random Variable Example (Y=X²)

- Let $Y = \Phi(X) = X^2$ *X* could denote the measurement error in a certain physical experiment, and *Y* would then be the square of the error.
- Note that $F_Y(y) = 0$ for $y \le 0$. For $y > 0$:

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y} - F_X(-\sqrt{y}) \end{aligned}$$

- And by differentiation the density of *Y* is:

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Functions of a Random Variable X=Std. Normal

- As a special case of Example 1, let *X* have the standard normal distribution [N(0,1)] so that:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty$$

- Then:

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left(\frac{1}{\sqrt{2\pi}} e^{-y/2} + \frac{1}{\sqrt{2\pi}} e^{-y/2}\right), & y > 0, \\ 0, & y \le 0, \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, & y > 0, \\ 0, & y \le 0. \end{cases}$$

Functions of a Random Variable X=Std. Normal (cont.)

- We conclude that Y has a gamma distribution with α = 1/2 and λ = 1/2.
- Since $GAM(\frac{1}{2}, n/2) = X_n^2$, it follows that if X is standard normal then $Y = X^2$ is a chi-square distributed with one degree of freedom.

Example, Simulation Application X=Uniform

- Let X be uniformly distributed on (0,1).
- Show $Y = -\lambda^{-1}$ 1n(1-X) has an exponential distribution with parameter $\lambda > 0$.
- Y is a nonnegative random variable implying $F_Y(y) = 0$ for $y \leq 0$.
- For $y > 0$:

$$F_Y(y) = P(Y \leq y) = P[-\lambda^{-1} 1n(1-X) \leq y]$$
$$= P[1n(1-X) \geq -\lambda y]$$
$$= P[(1-X) \geq e^{-\lambda y}]$$

since e^x is an increasing function of x,

$$= P(X \leq 1 - e^{-\lambda y})$$
$$= F_X(1 - e^{-\lambda y}$$

Example (cont.), Use in Simulation

- Since X is uniform over (0, 1), $F_X(x) = x, 0 \geq x \geq 1.$
- Thus, $F_Y(y) = 1 - e^{-\lambda y}$
- Y is exponentially distributed with parameter λ.
- This fact can be used in Monte Carlo simulation. In such simulation programs, it is important to be able to generate values of variables with known distribution functions (random deviates).
- Most computer systems provide built-in functions to generate random deviates from the uniform distribution over (0, 1), say u.
- To generate a random deviate, y, of an exponentially distributed random variable Y with the parameter λ, then from this example it follows that $y = -\lambda^{-1} 1n(1-u)$.

Theorem (generalize Ex. 2, 3)

- Let X be a continuous random variable with density f_X that is nonzero on a subset I of real numbers (i. e., $f_X(x) > 0, x \in I$ and $f_X(x) = 0, x \notin 1.)$
- Let Φ be a differentiable monotone function whose domain is I and whose range is the set of real numbers.
- Then $Y = \Phi(X)$ is a continuous random variable with the density, $f_Y(y)$ given by:

$$f_Y(y) = \begin{cases} f_X[\Phi^{-1}(y)][|(\Phi^{-1})'(y)|], & y \in \Phi(I) \\ 0, & \text{otherwise,} \end{cases}$$

- Where Φ^{-1} is the uniquely defined inverse of Φ and $(\Phi^{-1})'$ is the derivative of the inverse function.

Theorem (cont.)

- Proof:
  - We prove the theorem assuming that $\Phi(x)$ is an increasing function of x. The proof for the other case follows in a similar way.

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P[\Phi(X) \le y] \\ &= P[X \le \Phi^{-1}(y)], \quad \text{since } \Phi \text{ is monotone increasing} \\ &= F_X[\Phi^{-1}(y)] \end{aligned}$$

- Taking derivatives and using the chain rule, we get the required result.

Example 1, Simulation Issues

- Now let ϕ be the distribution function, F, of a random variable, X, with density f.
- Applying the theorem:

$$F_Y(y) = F_X(\phi^{-1}(y))$$

$$Y = F(X) \text{ and } F_Y(y) = F_X(F_X^{-1}(y)) = y.$$

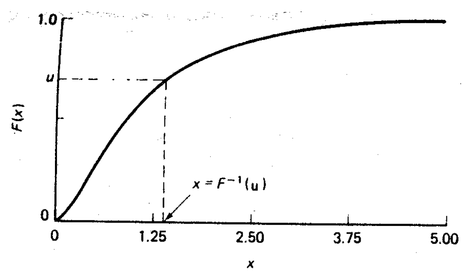
- Therefore, the random variable Y=F(X) has the density given by:

$$f_Y(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & otherwise. \end{cases}$$

- In other words, if X is a continuous random variable with CDF of F, then the new random variable Y=F(X) is uniformly distributed over the interval (0,1).

Example 1, Simulation Issues (cont.)

- This idea can be used to generate a random deviate x of X by:
  1. Generating a random number u from a uniform distribution of (0,1) and then
  2. Using the relation $x=F^{-1}(u)$ as illustrated in the figure below.
- The generation of the (0,1) random number can be done easily.
- Question is whether $x=F^{-1}(u)$ can be expressed in a closed mathematical form. This is possible for distributions such as the exponential; for distributions such as the normal, we must use other techniques.



Example 2: Log Normal Distribution

- Let X be normally distributed and consider $Y = e^X$
- Since:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

and $\Phi^{-1}(y) = 1n(y)$ implies that $[\Phi^{-1}]'(y) = \frac{1}{y}$

Example 2: Log Normal Distribution (cont.)

- Then, using the theorem, the density of Y is:

$$f_Y(y) = \frac{f(\ln y)}{y}$$
$$= \frac{1}{\sigma y\sqrt{2\pi}} \exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right], \qquad y > 0$$

- The random variable Y is said to have log-normal distribution.
- Another form of the central limit theorem, which states that the product of n mutually independent random variables has a log-normal distribution in the limit $n \to \infty$.

Example 3

- Consider a SCSI disk drive. Assume that the number of tracks N to be traversed between two disk requests is a normally distributed random variable with mean $\mu = n/3$ and $\sigma^2 = n^2/18$ where n is the total number of tracks. The seek time T is a random variable related to the seek distance N by:

$$T = a + bN$$

- Assume that experimental data show $a = 45$, $b = 0.43$, and $n = 200$. Determine the pdf of T. Recalling that T is a nonnegative random variable whereas the normal model allows negative values, make appropriate assumptions.

Example 3 (cont.)

- We use the theorem:
- If Φ be a differentiable monotone function whose domain is *I* and whose range is the set of real numbers.
- Then $Y = \Phi(X)$ is a continuous random variable with the density, $f_Y(y)$ given by:

$$f_Y(y) = \begin{cases} f_X[\Phi^{-1}(y)][\left|(\Phi^{-1})'(y)\right|], & y \in \Phi(I) \\ 0, & \text{otherwise,} \end{cases}$$

Example 3 (cont)

- If $\Phi(y)$ is linear, i.e., Y = cX+d.
- Then, use of the theorem yields:

$$f_Y(y) = \begin{cases} \frac{1}{|c|} f_X\left\{\frac{y-d}{c}\right\}, & y \in cI + d, \\ 0, & \text{otherwise,} \end{cases}$$

where I is the interval over which f(x) $\neq$ 0.

Example 3 (cont.)

- Since T must be positive, we use *truncated normal density*;

$$f(x) = \begin{cases} 0 & x < 0 \\ \dfrac{1}{\alpha\sigma\sqrt{2\pi}} \exp\left[\dfrac{-(x-\mu)^2}{2\sigma^2}\right] & x \geq 0 \end{cases}$$

- Where:

$$\alpha = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(t-\mu)^2}{2\sigma^2}\right] dt$$

Example 3 (cont.)

- In our case T will be normally distributed with

$$\mu_T = a + b\frac{n}{3} = 45 + 0.43\frac{200}{3} = 73.67$$

$$\sigma_T^2 = b^2\sigma^2 = (0.43)^2 \frac{n^2}{18} = 410.89$$

Example 3 (cont.)

- We show that α is close to 1:

$$\alpha = \int_0^\infty \frac{1}{\sigma_T \sqrt{2\pi}} \exp\left[\frac{-(t-\mu_T)^2}{2\sigma_T^2}\right] dt =$$

$$1 - F_Z\left(\frac{0-\mu_T}{\sigma_T}\right) = 1 - F_Z\left(\frac{-73.67}{20.27}\right) =$$

$$1-(1-F_Z(3.63)) = F_Z(3.63) = 0.99986$$