Lecture 3: More on Random Variables

ELE 525: Random Processes in Information Systems

Hisashi Kobayashi

Department of Electrical Engineering
Princeton University
September 23, 2013

Textbook: Hisashi Kobayashi, Brian L. Mark and William Turin, ***Probability, Random Processes and Statistical Analysis*** (Cambridge University Press, 2012)

Suppose that variables X and Y are *independent*, i.e.,

$$p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j), \quad \text{for all} \quad x_i, y_j. \tag{3.52}$$

Then

$$E[XY] = \sum_i \sum_j x_i y_j p_{XY}(x_i, y_j) = \left(\sum_i x_i p_X(x_i)\right)\left(\sum_j y_j p_Y(y_j)\right) = \mu_X \mu_Y, \tag{3.53}$$

Theorem 3.1 (Expectation of the product of independent random variables). *If random variables X and Y are* statistically independent *with finite expectations μ_X and μ_Y, their product XY is a random variable with expectation $\mu_X \mu_Y$:*

$$E[XY] = E[X]E[Y] = \mu_X \mu_Y. \tag{3.54}$$

□

Then in view of Definition 3.7 and (3.47), we readily have

Theorem 3.2 (Independence implies uncorrelatedness). *If X and Y are independent, then they are uncorrelated. However, the converse is not true.* □

Theorem 3.3 (Variance of sum of independent variables). *If X and Y are independent, the variance of $Z = X + Y$ is given by*

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2. \tag{3.55}$$

□

Theorem 3.4 (Sum of n random variables). *Let $X_1, X_2, \ldots, X_n$ be random variables with finite means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. Consider the sum variable*

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Then

$$E[S_n] = \mu_1 + \mu_2 + \ldots + \mu_n \tag{3.56}$$

$$\text{Var}[S_n] = \sum_{i=1}^{n} \sigma_i^2 + 2\sum_{i<j} \text{Cov}[X_i, X_j], \tag{3.57}$$

where the last sum extends over the $\frac{n(n-1)}{2}$ pairs (X_i, X_j) with $i < j$. In particular, if all X_i's are pairwise independent, then

$$\text{Var}[S_n] = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2. \tag{3.58}$$

3.3 Important Probability Distributions

3.3.1 Bernoulli distribution and binomial distribution

Equation (2.36) of p. 27 defines the *Bernoulli distribution* of *n* trials, and (2.38) defined the *binomial distribution*, i.e.,

$$P[X = k] = B(k; n, p) \triangleq \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n. \tag{3.62}$$

$$\mu_X = E[X] = np. \tag{3.64}$$

$$\sigma_X^2 = \mathrm{Var}[X] = npq. \tag{3.66}$$

See Problem 3.12 for an extension to the **multinomial distribution**.

3.3.2 Geometric distribution

- Consider a series of Bernoulli trials with $P[\{s\}]=p$.
  Let X=Number of trials until the first success.

$$\boxed{P[X=k]=q^{k-1}p, \quad k=1,2,\ldots,} \tag{3.68}$$

$$P[X>k]=q^k, \quad k=0,1,2,\ldots. \tag{3.69}$$

$$P[X>m+k|X>m]=\frac{P[X>m+k]}{P[X>m]}=\frac{q^{k+m}}{q^m}=q^k. \tag{3.70}$$

$$\boxed{E[X]=\frac{1}{p}.} \tag{3.74}$$

$$\boxed{\mathrm{Var}[X]=\frac{q}{p^2}.} \tag{3.75}$$

3,3,3 Poisson distribution

Let us consider the limiting case of the binomial distribution $B(k; n, p)$ of (2.38) when $n \to \infty$ and $p \to 0$, while keeping

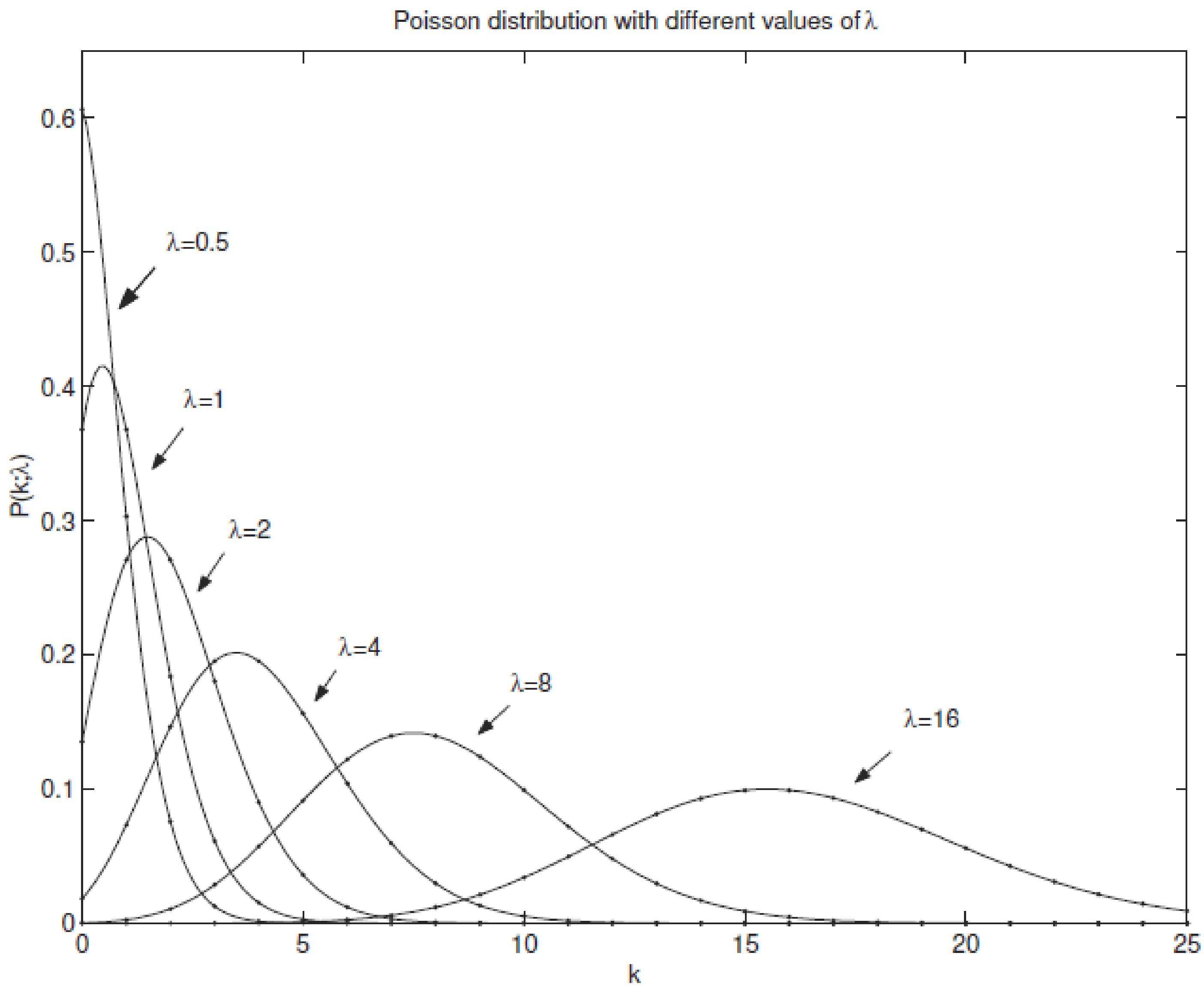
$$np = \lambda,$$

where λ is a fixed parameter. Then, as will be shown below, the limit distribution becomes

$$\boxed{P(k;\lambda) \triangleq \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, 2, \ldots,} \tag{3.77}$$

which is the *Poisson distribution* with *mean* λ.

$$\boxed{\begin{aligned} E[X] &= \lambda \\ E[X^2] &= \lambda^2 + \lambda \\ \mathrm{Var}[X] &= \lambda. \end{aligned}} \tag{3.85}$$


Poisson distribution with different values of λ
P(k;λ)
k
λ=0.5
λ=1
λ=2
λ=4
λ=8
λ=16
0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
15
20
25

3.3.4 Negative binomial (or Pascal) distribution

Let

$$Y_r \triangleq \text{Number of trials needed to achieve } r \text{ successes.} \quad (3.92)$$

Then we readily find

$$\begin{aligned} P[Y_r = k] &= P[r-1 \text{ successes in } k-1 \text{ trials and a success at the } k\text{-th trial}] \\ &= \binom{k-1}{r-1} p^{r-1} q^{k-r} p = \binom{k-1}{r-1} p^r q^{k-r}, \end{aligned} \quad (3.93)$$

where $q = 1 - p$. Thus, we have

$$\boxed{P[Y_r = k] = \binom{k-1}{r-1} p^r q^{k-r}, \quad k \geq r.} \quad (3.94)$$

Using the identity

$$\begin{aligned} \binom{-n}{i} &= \frac{(-n)!}{i!(-n-i)!} = (-1)^i \frac{n(n+1)\cdots(n+i-1)}{i!} \\ &= (-1)^i \binom{n+i-1}{i}, \end{aligned} \quad (3.97)$$

we may write (3.94) as

$$\boxed{P[Y_r = k] = \binom{-r}{k-r} p^r (-1)^{k-r} (1-p)^{k-r} = \binom{-r}{k-r} p^r (-q)^{k-r}, \quad k \geq r,} \quad (3.98)$$

$$\boxed{E[X] = \frac{r}{p}.} \quad (3.104)$$

$$\boxed{\text{Var}[Y_r] = \frac{rq}{p^2}.} \quad (3.106)$$

3.4.1 Shifted negative binomial distribution

Let

Z_r=Number of *failures* needed to achieve r successes,

which is related to the Y_r simply by

$$Z_r = Y_r - r. \quad (3.107)$$

$$f(k; r, p) \triangleq P[Z_r = k] = \binom{r+k-1}{r-1} p^r q^k, \quad k = 0, 1, 2, \ldots. \quad (3.108)$$

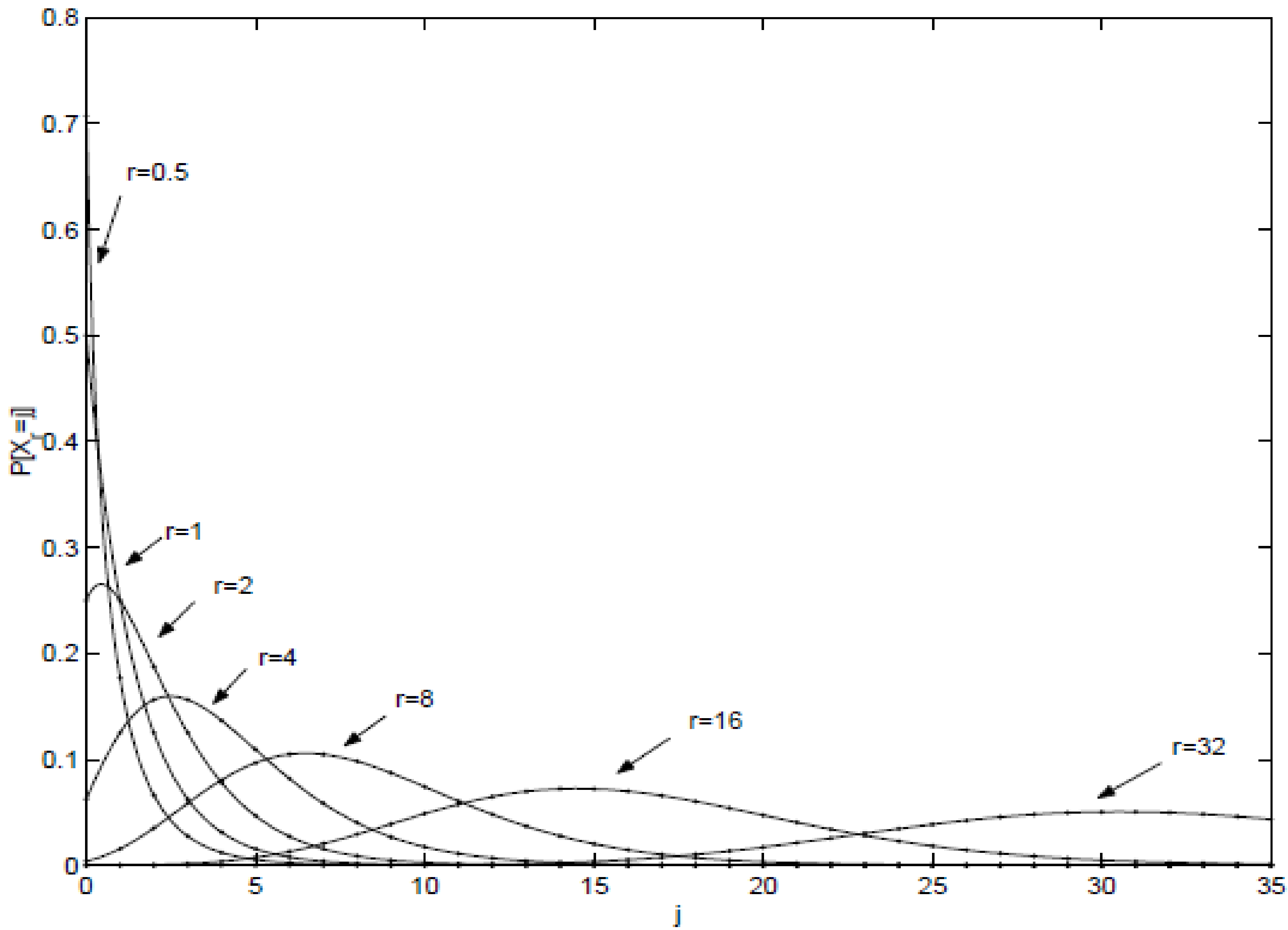


Figure 3.6 The *shifted* negative binomial distribution (3.108) or (3.109), i.e., the probability distribution of the number of *failures* needed to achieve r successes in Bernoulli trials with $p = 0.5$ and for $r = 0.5, 1, 2, 4, 8, 16, 32$.

4 Continuous Random Variables

4.1 Continuous Random Variables

- **Continuous random variable :** A random variable is called a **continuous random variable**, if its range is a continuum or, equivalently, if its distribution function is everywhere continuous.

4.1.1 Distribution function and probability density function

- The derivation of the distribution function $F_X(x)$, if it exists, is called the **probability density function (PDF)** of X:

$$f_X(x) = \frac{dF_X(x)}{dx}, \tag{4.3}$$

- In differential notation, we write formally

$$f_X(x)\,dx = dF_X(x) = P[x < X \leq x + dx]. \tag{4.4}$$

Therefore,

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du. \tag{4.5}$$

❖ Properties of the PDF:

$$f_X(x) \geq 0. \tag{4.6}$$

$$P[a < X \leq b] = \int_{a}^{b} f_X(x)\, dx. \tag{4.7}$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1. \tag{4.8}$$

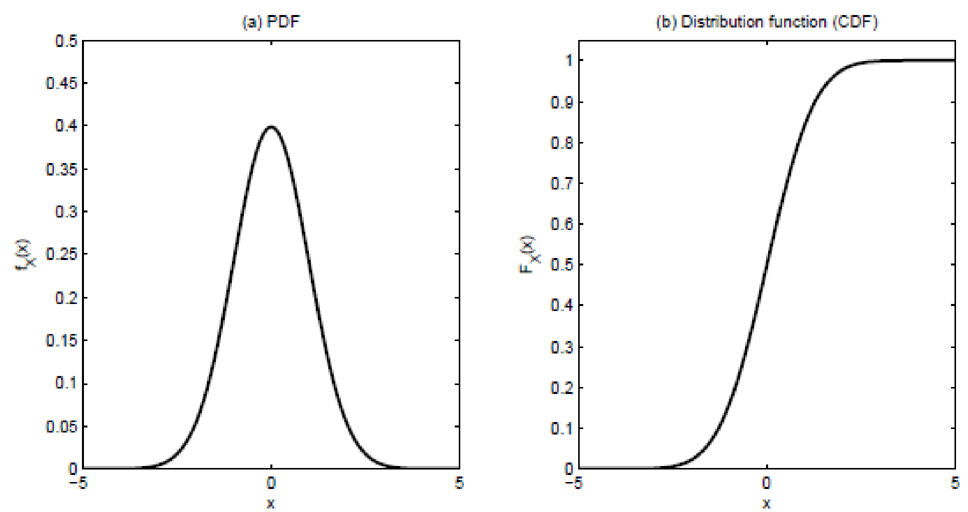


Figure 4.1 (a) The probability density function (PDF) and (b) the distribution function of a unit normal random variable.

4.1.2 Expectation, moments, central moments, and variance

The expectation:

$$\boxed{\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx,} \tag{4.9}$$

$$\begin{aligned} \mu_X &= \int_0^{\infty} x f_X(x)\, dx + \int_{-\infty}^{0} x f_X(x)\, dx \\ &= \int_0^{\infty} [1 - F_X(x)]\, dx - \int_{-\infty}^{0} F_X(x)\, dx. \end{aligned} \tag{4.10}$$

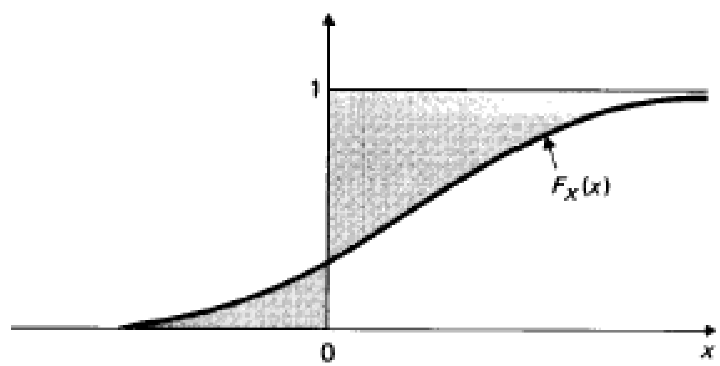


Figure 4.2 The expectation of the random variable X is the difference of the shaded regions.

❖ If *X* is a **nonnegative** RV,

$$\mu_X = \int_0^\infty F_X^c(x)\,dx, \tag{4.11}$$

where

$$F_X^c(x) = 1 - F_X(x), \quad x \geq 0 \tag{4.12}$$

Is the **complementary distribution function** or the **survivor function** of the RV *X*.

❖ The *m*th **moment** and *m*th **central moment** of a continuous RV *X*:

$$E[X^m] = \int_{-\infty}^{\infty} x^m f_x(x)\,dx, \tag{4.13}$$

$$E[(X - \mu_X)^m] = \int_{-\infty}^{\infty} (x - \mu_X)^m f_X(x)\,dx. \tag{4.14}$$

The variance:

$$\sigma_X^2 = \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\,dx$$
$$= E[X^2] - \mu_X^2. \tag{4.15}$$

❖ For a nonnegative RV *X*,

$$E[X^2] = 2\int_0^\infty x F_X^c(x)\,dx. \tag{4.16}$$

$$\sigma_X^2 = E[X^2] - \mu_X^2 = 2\int_0^\infty x F_X^c(x)\,dx - \mu_X^2. \tag{4.17}$$

❖ The notions and properties of **covariance** and **correlation coefficient** between continuous RVs are exactly the same as those for discrete RVs.

4.2 Important Continuous Random Variables and Their Distributions

4.2.1 Uniform distribution

$$F_X(x) = \begin{cases} 0, & \text{if } x < a, \\ \frac{x-a}{b-a}, & \text{if } a \le x \le b, \\ 1, & \text{if } x > b. \end{cases} \tag{4.18}$$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b, \\ 0, & \text{elsewhere}. \end{cases} \tag{4.19}$$

$$\mu_X = \frac{b+a}{2}, \tag{4.20}$$

$$E[X^2] = \frac{b^2+ba+a^2}{3}, \tag{4.21}$$

$$\sigma_X^2 = \frac{(b-a)^2}{12}. \tag{4.22}$$

❖ The unit uniform RV *U* defined over the unit interval [0, 1] has

$$\begin{aligned} F_U(u) &= u, \;\; 0 \le u \le 1, \\ f_U(u) &= 1, \;\; 0 \le u \le 1,. \end{aligned} \tag{4.23}$$

4.2.2 Exponential distribution

The **exponential distribution function** with **rate** λ

$$F_X(x) = 1 - e^{-\lambda x}, \ \ x \geq 0, \tag{4.24}$$

and the density function

$$f_X(x) = \lambda e^{-\lambda x}, \ \ x \ \geq 0. \tag{4.25}$$

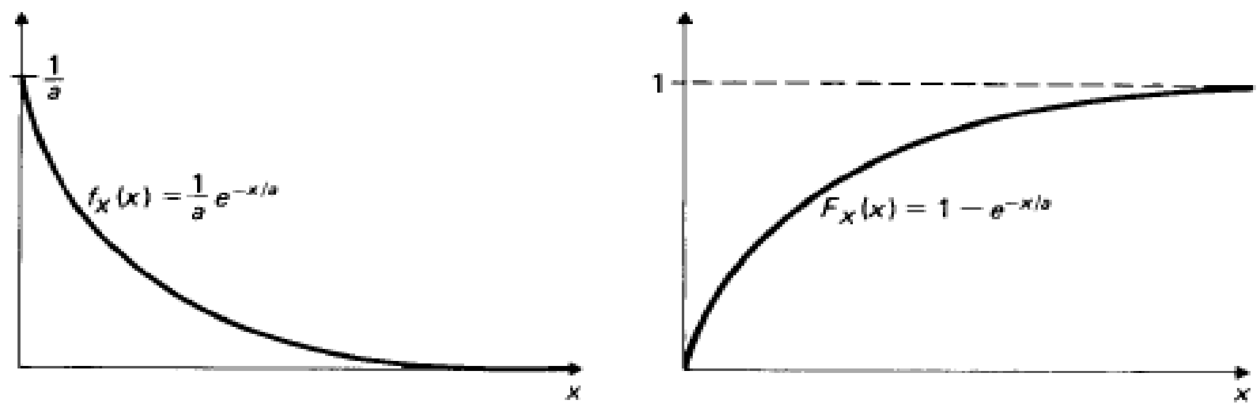


Figure 4.3 (a) The PDF and (b) the distribution function of an exponential RV.

Note: a=1/λ is the mean of the distribution

❖ The **memoryless property** of the exponential distribution.

$$P[X > t + s | X > s] = \frac{P[X > t + s]}{P[X > s]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P[X > t]. \quad (4.26)$$

❖ The "waiting time paradox of the **Poisson process**."

❖ The mean and variance:

$$\mu_X = \int_0^\infty F_X^c(x)\, dx = \int_0^\infty e^{-\lambda x}\, dx = \frac{1}{\lambda}. \quad (4.27)$$

$$\sigma_X^2 = 2\int_0^\infty x e^{-\lambda x} - \mu_X^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \quad (4.28)$$

❖ The coefficient of variation

$$c_X = \frac{\sigma_X}{\mu_X}, \quad (4.29)$$

is equal to unity.

4.2.3 Gamma distribution

The PDF of the gamma distribution with parameters (λ,β):

$$f_Y(y;\lambda,\beta) \triangleq \frac{\lambda(\lambda y)^{\beta-1}e^{-\lambda y}}{\Gamma(\beta)}, \quad y \geq 0; \beta, \lambda > 0, \tag{4.30}$$

where $\Gamma(\beta)$ is the gamma function

$$\Gamma(\beta) = \int_0^\infty t^{\beta-1}e^{-t}\,dt = (\beta-1)\Gamma(\beta-1), \tag{4.31}$$

- The case β=1 reduces to the exponential distribution with rate λ.
- The case λ=1 is called the **standard gamma distribution**:

$$f_X(x;\beta) \triangleq \frac{x^{\beta-1}e^{-x}}{\Gamma(\beta)}, \quad x \geq 0; \beta > 0. \tag{4.32}$$

$$E[X] = \frac{1}{\Gamma(\beta)}\int_0^\infty \lambda x e^{-x} x^{\beta-1}\,dx = \frac{\Gamma(\beta+1)}{\Gamma(\beta)} = \beta, \tag{4.36}$$

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2 = \beta. \tag{4.38}$$

❖ For the RV *Y* defined by (4.30), we find

$$E[Y] = \frac{\beta}{\lambda}, \quad E[Y^2] = \frac{\beta(\beta+1)}{\lambda^2}, \quad \text{and} \quad \text{Var}[Y] = \frac{\beta}{\lambda^2}. \tag{4.39}$$

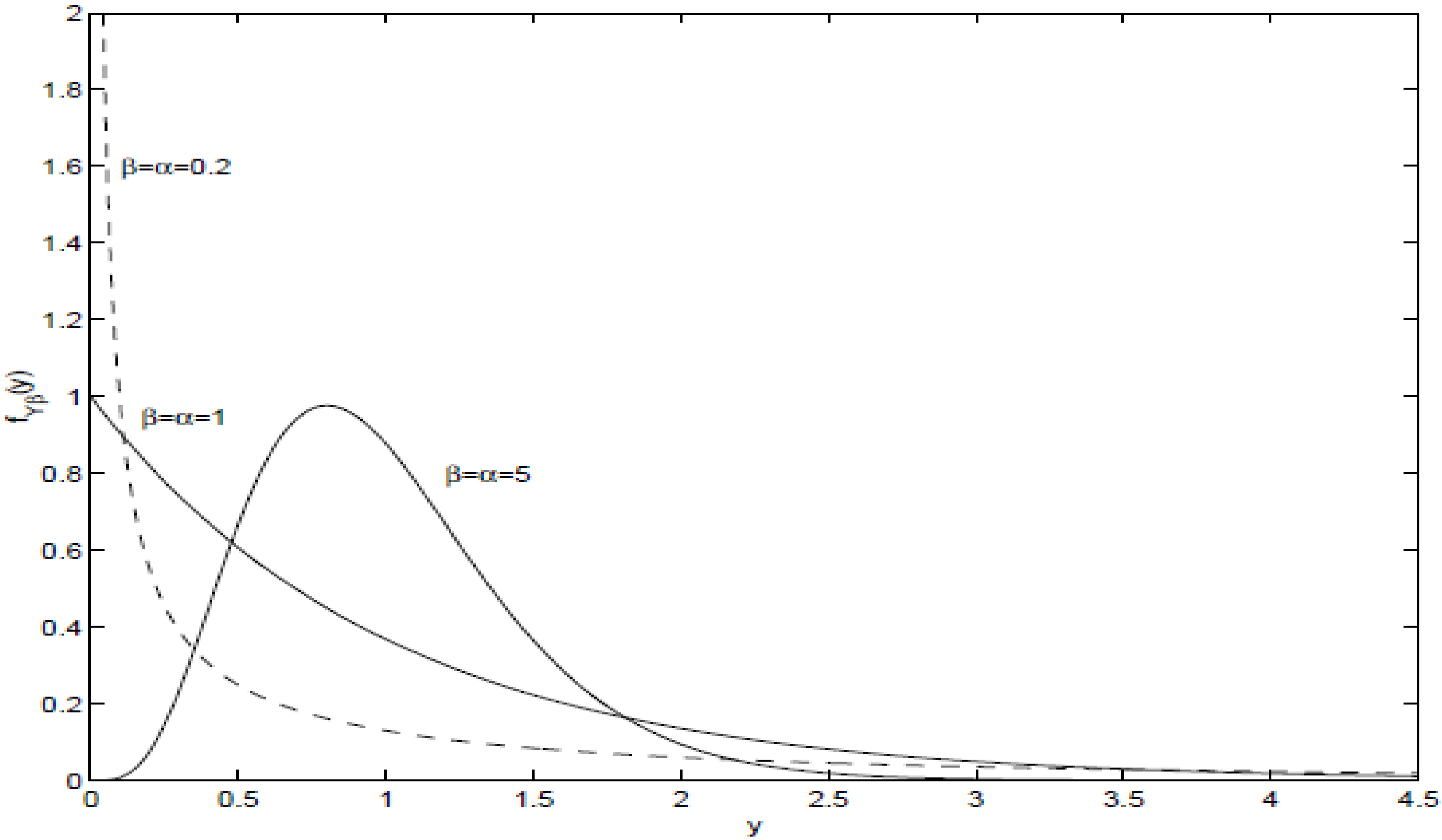


Figure 4.4 Gamma distributions for different parameters (λ, β), where the mean is kept constant $(= 1)$, i.e., $\beta/\lambda = 1$.

Note: α should be λ

- For $\beta < 1$ (the dashed curve in Figure 4.4) , the gamma distribution may fit a distribution with **a long tail**. But it goes to infinity as *y* --> 0.
- For $\beta = 1$, the gamma distribution reduces to the **exponential distribution** with rate λ.
- For $\beta > 1$, the PDF starts from the origin, and takes a single maximum (i.e., its mode) at $x = \beta - 1$ (i.e., at $y = (\beta - 1)/\lambda$) and decreases towards zero as *x* (or *y*) --> ∞.
- If $\beta = r$, a positive integer, the RV *Y* is equivalent to the sum of *k* i.i.d. exponential variables with mean $1/\lambda$. Hence $E[Y] = r/\lambda$. The distribution is known as the *r*-stage **Erlang distribution** (4.163) of Problem 4.11.

$$F_{Y_r}(y) = 1 - e^{-\lambda y} \sum_{j=0}^{r-1} \frac{(\lambda y)^j}{j!}, \quad y \geq 0, \tag{4.163}$$

- Thus, the gamma variable can be viewed as the sum of β i.i.d. exponential random variables. Then by virtue of the **central limit theorem (CLT)** the gamma distribution should approach the **normal distribution** as β becomes large, with mean $(\beta - 1)/\lambda \approx \beta/\lambda$ and variance β/λ^2.
- If $\beta = n/2$ for integer n in the standard gamma distribution (4.32), it become the **chi-squared distribution** with *n* degrees of freedom. (see Section 7.4).

❖ Application examples of the gamma distribution

- waiting time in queueing systems
- the lifetime of devices as in reliability theory
- the load on webservers
- the distribution of rainfall in climatology
- the distribution of insurance claims in finance service
- widely used as a **conjugate prior** (see pp.97-102) in Bayesian statistics. It is the conjugate prior for the precision (i.e. inverse of the variance) of a normal distribution. It is also the conjugate prior for the exponential distribution.

4.2.4 Normal (or Gaussian) distribution

The **normal** (or **Gaussian**) distribution with *mean* μ and *variance* σ^2:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (4.40)$$

This distribution is often denoted by $N(\mu, \sigma^2)$.

❖ Define $U=(X-\mu)/\sigma$. Then the variable has the distribution $N(0,1)$, whose PDF is

$$\phi(u) \triangleq \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\}, \tag{4.41}$$

and the distribution function is denoted by $\Phi(u)$:

$$\Phi(u) = \int_{-\infty}^{u} \phi(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp\left\{-\frac{t^2}{2}\right\}\, dt \tag{4.46}$$

1. For any $\infty < x < \infty$,

$$\Phi(-x) = 1 - \Phi(x). \tag{4.47}$$

2. For any $x > 0$,

$$\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3}\right) < 1 - \Phi(x) < \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \cdot \frac{1}{x}. \tag{4.48}$$

3. For large x,

$$1 - \Phi(x) \approx \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}x}, \quad x \gg 1. \tag{4.49}$$

4.2.4.1 Moments of the unit normal distribution

$$E[U^n] = 0 \quad (n \text{ odd}). \tag{4.53}$$

When $n \geq 2$ is even, we obtain

$$E[U^n] = 1 \cdot 3 \cdot 5 \cdots (n-1) \quad (n \text{ even}), \tag{4.54}$$

In particular,

$$E[U] = 0 \text{ and } \sigma_U^2 = E[U^2] = 1. \tag{4.55}$$

For $X \sim N(0,1)$, we can write

$$X = \sigma U + \mu. \tag{4.56}$$

- The reproductive property of the normal variables.
  Let X_i , i=1, 2, …, n be independent RVs having distributions $N(\mu_i, \sigma_i^2)$, and let

$$Y = \sum_{i=1}^{n} a_i X_i, \tag{4.58}$$

where a_i are real constants. Then the distribution of Y is also normal:

$$N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$