Chapter 2 Univariate Probability Distribution

- Goldberger, Ch 2
- Wackerly et al. chapter 3, 4

  Yale note, chapter 4

- Random variable
- Discrete random variable: Probability mass function,

  Cumulative distribution function

  Uniform, Bernoulli, Binomial, Mulnomial, Poisson
- Continuous random variable: Probability density function

  Median, Percentile

  Uniform, Exponential, Normal

► Functions of Random Variable, Mixed Distribution

1. Introduction

◎ Statistics is concerned with controlled experiment for which the outcome cannot be predicted with certainty. Such experiments are called *random experiment*.

• *Sample space* (S) : set of all possible outcomes.

(Ex) If rat is selected from a cage at random and the sex is determined, $S = \{male,\ female\}$.

(Ex) If survey is conducted about the number of children in the household, $S = \{0, 1, 2, \cdots\}$.

(Ex) The weight of randomly selected candy bar of a given brand is w. $S = \{w \mid w > 0\}$.

◎ We can convert outcomes such as these into numbers by defining a random variables, denoted by X, which is a function that maps outcomes in S onto the real line: $X(s)=x, \quad s \in S,\ x \in R \quad (X : S \Rightarrow R)$

• (Definition) *Random variable*: real valued function that assigns a number to each sample point (outcome) in the sample space of the experiment; $S \rightarrow R$.

(Example) $X(F)=0, X(M)=1$, that is, $X=\begin{cases} 1 \text{ if male} \\ 0 \text{ if female} \end{cases}$.

► *Discrete r.v.* : which value is finite or countably infinite.

(Example) X=the number of chips in the randomly chosen cookie.

► *Continuous r.v.* : which value is at one interval of real line

(Example) X=number of calories in the randomly chosen cookie.

2. Discrete Random Variable

(Definition) For a discrete random variable, X, a *probability distribution (probability mass function; pmf)* is defined to be the function $f(x)$ such that for any real number x, which a value X can take,

$f(x) = P(X = x)$; probability mass function (p.m.f.).

Consequently,

① $0 \le f(x) \le 1$,

② $\sum_{x \in S} f(x) = 1$,

③ $P(x \in A) = \sum_{x \in A} f(x)$.

(Note)

$X, Y, \cdots$: random variable.

$x, y, \cdots$: a specified value.

◎ *<u>Cumulative distribution function</u>(c.d.f., or <u>distribution function</u>)*

$$F(x) = P(X \le x) = \sum_{t \le x} f(t)$$

① $F(-\infty) = 0$

② $F(\infty) = 1$

③ If $a < x < b$, then $F(a) \le F(b)$ for any real number a, b.

(Example) Tossing a coin example, $f(x)$, $F(x)$.

(Note) $F(x)$ is defined for all real numbers, thus, $F(1.5) = P(X \le 1.5) = 3/4$, even though $P(X = 1.5) = 0$.

• For discrete random variable, the c.d.f. is a step function and is right-continuous. If the range of a random variable X consists of the values $x_1 < x_2 < \cdots < x_n$,

① $F(x_1) = f(x_1)$

② $f(x_i) = F(x_i) - F(x_{i-1})$ for $i = 2, 3, \cdots, n$ $\Rightarrow$ $F(x_i) = F(x_{i-1}) + f(x_i)$

3. Special Discrete Distribution

(1) Discrete uniform distribution

- Equal probability on each of the points in its space.

(Example) Rolling a fair die

$f(x) = \frac{1}{6}$ **for** $x = 1, 2, \cdots, 6$.

- In general, if there are n possible outcomes, $f(x) = \frac{1}{n}$ for $x = 1, 2, \cdots, n$.

► $E(X) = \frac{m+1}{2}$, $E(X^2) = \frac{(m+1)(2m+1)}{6}$, $V(X) = \frac{m^2 - 1}{12}$.

(2) <u>Bernoulli distribution</u>

◎ A *<u>Bernoulli experiment</u>* is a random experiment, the outcome of which can be classified in one of two mutually exclusive ways.

(Example) Success or Fail, Male or Female, Head or Tail.

• Let X be random variable associated with a Bernoulli trial.
We define $X(success) = 1$, $X(fail) = 0$.

• $P(X = s) = p$, $P(x = f) = 1 - P(X = s) = 1 - p$

$f(x)$

$\mu = E(X) = p$, $\sigma^2 = V(X) = p(1 - p)$.

(3) Binomial distribution

◎ In a sequence of n (independent) Bernoulli trials, we are interested in total number of successes and not in the order of their occurrence.

(If interested in the order, then negative Binomial distribution).

• Let X denote the number of successes, x successes with p^x, then $(n-x)$ fails with $(1-p)^{n-x}$.

► The number of ways of selecting x successes in n trials: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

☞ $f(x)$

$\Rightarrow X \sim B(n,p)$.

(Ex) $P(H)=\frac{2}{3}$, $P(T)=\frac{1}{3}$.

X =# of H in 10 trials$\sim B\left(10, \frac{2}{3}\right)$.

- Bernoulli vs. Binomial:

Let $R=\begin{cases} 1 & \textbf{if success} \\ 0 & \textbf{if fail} \end{cases}$, with $P(R=1)=p,\ P(R=0)=1-p$,

X =# of successes in n trials$=\sum_{i=1}^{n} R_i$.

Then, $E(X)=np, V(X)=np(1-p)$.

(4) Multinomial distribution

◎ Each trial has more than two possible outcomes, with probabilities $p_1,\ p_2,\cdots,\ p_K$

such that $\sum_{i=1}^{k} p_i = 1$,

$$\Rightarrow\ f(x_1, x_2, \cdots, x_k; n, p_1, p_2, \cdots, p_k) = \binom{n}{x_1, x_2, \cdots, x_k} p_1^{x_1}\ p_2^{x_2} \cdots p_k^{x_k}$$

for $x_1 + x_2 + \cdots + x_k = n$ and $\sum_{i=1}^{k} p_i = 1$ where $\binom{n}{x_1, x_2, \cdots, x_k} = \dfrac{n!}{x_1!x_2!\cdots x_k!}$

(Example) Channel 6 has 20% of the viewing audience, Channel 9 has 30% of the viewing audience, and Channel 11 has 50% of the viewing audience. Among 8 randomly selected, 5 for channel 11, 2 for channel 9 and 1 for channel 6.

☞ $f(x_1 = 5, x_2 = 2, x_3 = 1)$

(5) Poisson Distribution

- $X \sim Poisson(\lambda), \;\; x = 0, 1, 2, \cdots$

$$P(X = x) = f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}.$$

Given time interval or space, random variable takes 0, 1, 2, …

① Integer between 0 and ∞.

② # of occurrences in one unit (time or space) is independent of that in any other unit.

③ Probability of occurrences in any unit is proportional to the size of the unit.

④ Probability of two or more occurrences in a sufficiently short interval is zero.

(Example) # of accidents, # of patents, # of sales.

(Example) At Inchon airport, in average two airplanes arrive at the airport per minutes.

$P(X = 8 | 8:00 - 8:03) = ?$

During 8:00-8:03, in average six airplanes arrives, so $\lambda = 6$.

Since $f(x) = \dfrac{e^{-6} 6^x}{x!}$, so $f(8) = \dfrac{e^{-6} 6^8}{8!} = 0.1033$.

☞ $E(X)$, $V(X)$

3. Continuous Random Variable

◎ Suppose that X can take any possible values $a \le X \le b$. Suppose the possible values of X are uncountable, the probability that X takes any particular value is zero(that is, zero probability mass on any given point of the support).

That is, $P(X = x) = 0$

Thus, the *probability density function*(*p.d.f.*) of a continuous random variable, $f(x)$ will NOT give the probability that X takes the value x. Instead, the area under $f(x)$ gives probability for corresponding interval.

- $f(x)$ is a probability density function if

① $f(x) \geq 0$ for all x.

② $\int_{-\infty}^{\infty} f(x)dx = 1$.

③ For any a, b with $-\infty < a < b < \infty$, $P(a \leq X \leq b) = \int_a^b f(x)dx$.

(Cf) $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

• The *<u>cumulative distribution function</u>(c.d.f.)* for a continuous random variable X is given by $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$ for $-\infty < x < \infty$, where $f(t)$ is the value of the p.d.f. at t.

Note that

① $F(-\infty) = 0$

② $F(\infty) = 1$

③ $F(a) \le F(b)$ when $a < b$

④ Furthermore, $P(a \le X \le b) = F(b) - F(a)$ and $f(x) = \frac{dF(x)}{dx} = F'(x)$,

when the derivative exists.

(Example) Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} 3e^{-3x} & x > 0 \\ 0 & otherwise \end{cases} \quad \Rightarrow$$ exponential distribution

To verify whether $f(x)$ is proper p.d.f.

① $f(x) \geq 0$ for all x,

② $\int_{-\infty}^{\infty} f(x)dx = 1$

Note that a density may exceed 1 unlike a probability,

- $P(0.5 \leq X \leq 1) = ??$

☞ c.d.f. $F(x)$

☞ $P(0.5 \leq X \leq 1)$

◎ *<u>Percentiles</u>*(of a continuous distribution)

(Definition) Let p be a number between 0 and 1. The ($100p$)th *<u>percentile</u>* of the distribution of a continous random variable X, $\eta(p)$ is defined by the smallest value such that

$$p = P(X \le \eta(p)) = F\left(\eta(p)\right)$$

Roughly, $\eta(p) = F^{-1}(p)$.

► Some prefer to call $\eta(p)$ pth <u>quantile</u>.

► For discrete distribution, the ($100p$)th *<u>percentile</u>* of the distribution $\eta(p)$ is the smallest value such that $p \le F\left(\eta(p)\right) = \int_{-\infty}^{\eta(p)} f(x)dx$.

(Example) $f(x) = \begin{cases} \dfrac{3}{2}(1-x^2) & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

Then, $F(x) = \dfrac{3}{2}\left(x - \dfrac{x^3}{3} \right)$.

$(100p)$th percentile satisfies $p = F(\eta(p)) = \dfrac{3}{2}\left(\eta(p) - \dfrac{\eta(p)^3}{3} \right)$.

$\Rightarrow \ \eta(p)^3 - 3\eta(p) + 2p = 0$.

If $p = 0.5$, then $\eta^3 - 3\eta + 1 = 0 \quad \Rightarrow \eta(0.5) = 0.347$

(Definition) The ***median*** of continuous distribution, m, is 50th percentile, so m satisfies $0.5 = F(m)$.

- Some special percentiles:

► Median: 50^{th} percentile.

► Quantiles: 25, 50, 75 percentiles.

► Deciles: 10, 20, 30, …, 80, 90 percentiles.

(Example) Percentiles in income distribution.

- **상하위 20% 소득격차 사상최대**

올해 2·4분기 소득하위 20% 계층(1분위)의 명목 월평균소득은 지난해 같은 기간보다 2.7% 감소한 반면 상위 20% 계층(5분위)의 명목 월평균소득은 2.2% 줄어드는 데 그쳤다. 이에 따라 상위 20% 계층의 소득을 하위 20% 계층의 소득으로 나눈 소득 5분위 배율이 7.29배로 높아지면서 2·4분기 기준으로 소득격차가 가장 크게 벌어진 것으로 나타났다.
<중략>

◇ 소득격차 더 벌어져 = 통계청이 28일 내놓은 '2·4분기 가계동향'에 따르면 전국 가구(2인 이상)의 명목 월평균소득은 329만9000원으로 지난해 같은 기간(330만2000원)에 비해 0.1% 감소했다. 전국 가구의 명목소득이 줄어든 것은 2003년 이후 처음이다. 소득하위 20% 계층의 명목 월평균소득은 90만2000원으로 1년 전보다 2.7% 감소한 반면 소득상위 20%의 월평균소득은 657만6000원으로 2.2% 줄었다.

이에 따라 소득 5분위 배율은 7.29배로 벌어져 2003년 관련 통계 작성 이후 최대치를 기록했다. 2·4분기 기준으로 소득 5분위 배율은 2006년 7.16배, 2007년 7.22배, 지난해 7.25배 등으로 매년 높아지는 추세를 보이고 있다.

4. Special Continuous Distribution

(1) Uniform distribution (Rectangular)

- $f(x) = \begin{cases} \dfrac{1}{b-a} & \textbf{for } a < x < b \\ 0 & \textbf{otherwise} \end{cases}$

$\Rightarrow$ $X \sim U(a,b)$

☞ $\mu = E(X) = \dfrac{a+b}{2}$

☞ $\sigma^2 = V(X) = \dfrac{(b-a)^2}{12}$

☞ $F(x)$

(2) Exponential distribution

- $f(x) = \begin{cases} \lambda e^{-\lambda x} & for\ x > 0 \\ 0 & elsewhere \end{cases}$ with $\lambda > 0$

☞ $F(x)$

- Exponential distribution is appropriate for length of time until a light bulb fails and duration of unemployment (or waiting time).

$\mu = E(X) = \dfrac{1}{\lambda}$, $\sigma^2 = V(X) = \dfrac{1}{\lambda^2}$

(3) Normal distribution(Goldberger ch. 7-1)

• The *normal distribution* was developed to model the bell-shaped distribution of many random variables based on homogenous populations.

- $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(X-\mu)^2}{2\sigma^2} \right), \quad -\infty < x < \infty$

  $\Rightarrow \quad X \sim N(\mu, \sigma^2)$

- $\mu = E(X), \quad \sigma^2 = V(X)$.

(Remark)

► μ is a location parameter.

⇒ If shifts the normal distribution.

► σ is a scale parameter.

⇒ It scales the distribution by some factor.

(Special case) $\mu = 0$, $\sigma^2 = 1$ $\Rightarrow$ *Standard normal distribution*.

- $Z \sim N(0,1) \;\Rightarrow\; \phi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, $-\infty < x < \infty$.

- If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim N(0,1)$.
- If $Z \sim N(0,1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

- A linear function of normal random variable has also normal distribution

► If $X \sim N(\mu, \sigma^2)$, then $Y = a + bX \sim N(a + b\mu,\; b^2\sigma^2)$.

- Shape of $\phi(z)$: a familiar bell-shaped curve.

① $\phi(-z)=\phi(z)$; symmetric about 0.

② The ordinate at zero is $\phi(0)=\dfrac{1}{\sqrt{2\pi}}=0.3989$.

③ $\phi'(z)=-z\phi(z)$.

④ $\phi''(z)=-\phi(z)(1-z^2)$; has inflection points at $z=\pm 1$.

- The c.d.f. is $P(Z \le z) = \int_{-\infty}^{z} \phi(t)dt \equiv \Phi(z)$; No closed form.

- Percentiles of $N(0,1)$:

  From N(0,1) table,

$\eta(0.995)$	$\eta(0.005)$
$\eta(0.975)$	$\eta(0.025)$
$\eta(0.95)$	$\eta(0.05)$

4. Functions of Random Variables

• Consider a random variable X and its p.f. $f(x)$.

We shall be interested in finding the probability function of $Y = u(X)$.

That is, given $X \sim f(x)$, $Y = u(X) \Rightarrow Y \sim g(y),\ g(y) = ?$

► We will illustrate for continuous random variables cases.

(1) Distribution function technique

- Given continuous random variable X and $Y = u(X)$,

$$G(y) = P(Y \le y) = P\big(u(X) \le y\big)$$

$$\Rightarrow g(y) = \frac{dG(y)}{dy}$$

(Example) Given $f(x)=\begin{cases} 6x(1-x) & 0<x<1 \\ 0 & otherwise \end{cases}$, find p.d.f. of $Y=X^3$.

☞ $G(y)$

☞ $g(y)$

(2) Change-of-variable technique(transformation technique)

(a) Discrete case

- We can obtain the probability distribution of the transformed variable by simple substitution.

(Example 1) Consider $X \sim B\left(4, \frac{1}{2}\right)$.

That is $B\left(4, \frac{1}{2}\right) = \binom{4}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{4-x} = \binom{4}{x}\left(\frac{1}{2}\right)^{4}$.

Thus,

x	0	1	2	3	4
f(x)	1/16	4/16	6/16	4/16	1/16

Find pmf of $Y = \frac{1}{1+X}$

x	0	1	2	3	4
y	1	1/2	1/3	1/4	1/5
f(y)	1/16	4/16	6/16	4/16	1/16

Since $f(x) = \binom{4}{x}\left(\frac{1}{2}\right)^4$ for x=0, 1, 2,3, 4,

$$g(\mathrm{y}) = \binom{4}{\frac{1}{y}-1}\left(\frac{1}{2}\right)^4$$ for y=1, 1/2, 1/3,1/4, 1/5.

(Example 2) In the same example, find the pmf of $Z=(X-2)^2$.

x	0	1	2	3	4
f(x)	1/16	4/16	6/16	4/16	1/16
z	4	1	0	1	4
h(z)	1/16	4/16	6/16	4/16	1/16

Therefore,

h(0)=f(2)=6/16,

h(1)=f(1)+f(3)=8/16

h(4)=f(0)+f()=2/16

and we find

z	0	1	4
h(z)	3/8	4/8	1/8

(b) Continuous case

(Theorem) If $Y = u(X)$ is differentiable and <u>either increasing or decreasing</u> (monotone) for all values with the range of X for which $f(x) \neq 0$ so that the inverse function $X = u^{-1}(Y) = w(Y)$ exists and is differentiable, then p.d.f. of Y is $g(y) = f\left[w(y)\right]\left|w'(y)\right|$.

(Note) $w'(y) = \dfrac{dx}{dy}$ is called “Jacobian” of the transformation.

(Proof)

$$G(y) = P(Y \le y) = P(u(X) \le y)$$

① monotone increasing case

$$P(u(X) \le y) = f\left[w(y)\right]w'(y)$$

② monotone decreasing case

$$P(u(X) \le y) = -f(w(y))w'(y)$$

①,② $\Rightarrow$ $g(y) = f\left[w(y)\right]\left|w'(y)\right|$

(Note) Convenient formula:

$$g(y) = f(x)\left|w'(y)\right|$$

$$\Rightarrow g(y) = f(x)\left|\frac{dx}{dy}\right|$$

$$\Rightarrow g(y)\left|dy\right| = f(x)\left|dx\right|$$

(Example) $f(x) = \begin{cases} e^{-x} & for\ x > 0 \\ 0 & elsewhere \end{cases}$, find p.d.f. of $Y = \sqrt{X}$.

Since $\sqrt{X}$ is differentiable and monotone increasing,

$X = w(y)$ ☞

☞ $g(y)$

(Example) Given $f(x) = \begin{cases} 6x(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$,

Find the probability function of $Y = X^3$.

(Answer)

$$g(y) = \begin{cases} 2(y^{-1/3} - 1) & 0 < y < 1 \\ 0 & elsewhere \end{cases}$$

(Example) $Y \sim N\left(0, \sigma^2\right)$.

Represent $P(y > T)$ in terms of $\phi(.)$ or $\Phi(.)$.

- Truncated distribution: $f(y \mid y > T) = \dfrac{1/\sigma \cdot \phi(y/\sigma)}{1 - \Phi(T/\sigma)}$

(Example) Log-normal distribution

Suppose $X \sim N\left(\mu, \sigma^2\right)$, then $Y = \exp(X)$ has **log-normal** distribution.

(Or Y has log-normal distribution, then $X = \ln(Y)$ has normal distribution).

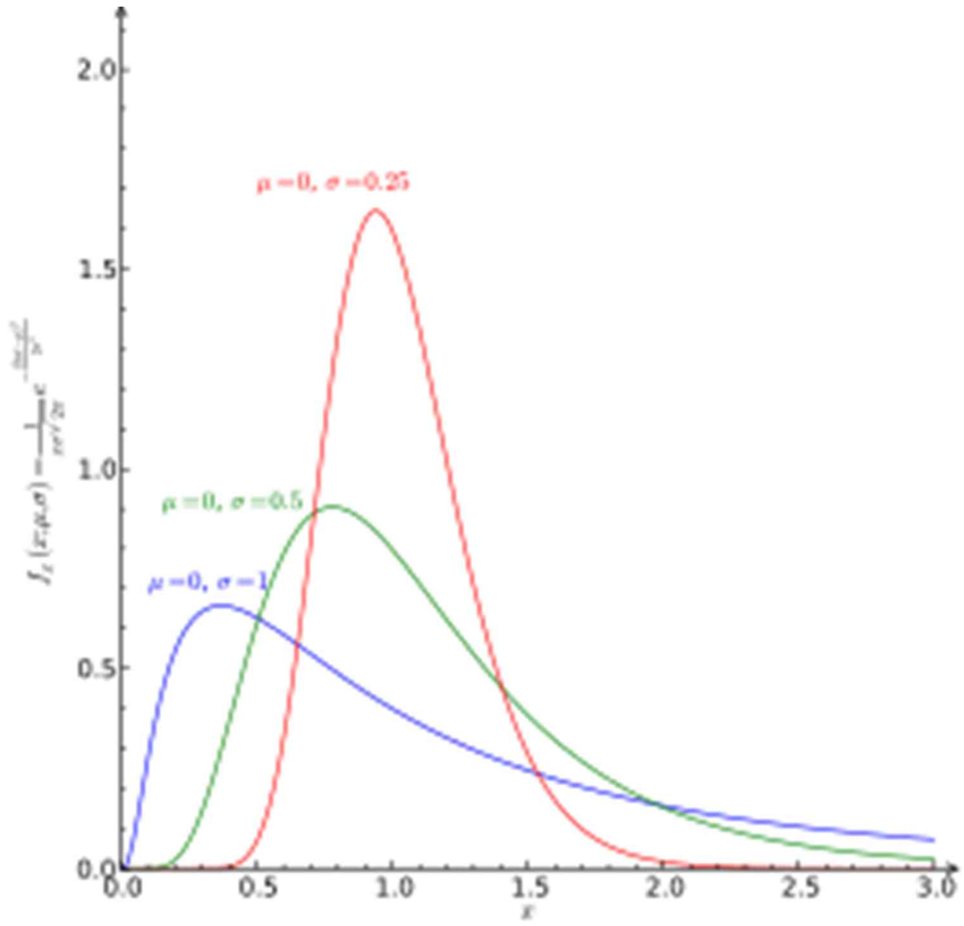
$$g(y) = \begin{cases} \dfrac{1}{y\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(\ln y - \mu)^2}{2\sigma^2} \right), & y > 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$E(Y) = e^{\mu + \frac{\sigma^2}{2}}, \quad V(Y) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$$

Median= $Median = \exp(\mu)$

► $P(Y \le y) = P(X \le \ln(y))$ since $\ln(y)$ is a monotone function.

(Example) When $\mu = 4, \quad \sigma^2 = 1$ in log-normal distribution,
find $P(Y \le 4)$ and $P(Y > 8)$.


μ=0, σ=0.25
μ=0, σ=0.5
μ=0, σ=1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x

- Monotone transformation preserves the ordering.
  - ► $M(Y) = u(M(\mathrm{X}))$ for linear or nonlinear function of $u(\cdot)$.
  - ► However, $E(Y) \neq u(E(X))$ in general.

  (special case) $E(Y) = u(E(X))$ if $u(\cdot)$ is linear function.

5. Mixed Distribution

► Most distributions are <u>either</u> discrete <u>or</u> continuous.

• Now, consider a distribution that is a mixture of a discrete distribution and a continuous distribution.

(Example) Y denotes the amount paid out by automobile insurance.

For many policies, Y=0 because the insured not involved in accidents.

For who do have accidents, the amount paid can be modeled with continuous distribution.

• A random variable Y that has some of its probability at discrete points and the remainder spread over intervals is said to have a *mixed distribution*.

$$F(y) = c_1 F_1(y) + c_2 F_2(y)$$

where $F_1(\mathrm{y})$ is a step distribution function,

$F_2(\mathrm{y})$ is a continuous distribution function,

c_1 is the accumulated probability of all discrete points.

$c_2 = 1 - c_1$ is the accumulated probability of all continuous portions.

(Example) Let Y denote the length of life of electronic components. These components frequently fail immediately with observed probability 1/4. If it does not fail immediately, the distribution for its length of life has exponential density function, $f(y) = e^{-y}, \quad \text{for} \quad y > 0$.

Find the distribution function for Y and evaluate $P(Y > 10)$.

(Solution)

$$F_1(y) = \begin{cases} 0, & y < 0 \\ 1, & y \ge 0 \end{cases} \qquad F_2(y) = \begin{cases} 0, & y < 0 \\ 1 - e^{-y}, & y \ge 0 \end{cases}$$

then, $F(y) = (1/4)F_1(y) + (3/4)F_2(y)$

$P(Y > 10) = 1 - F(10) = 1 - [(1/4) + (3/4)(1 - e^{-10})] = (3/4)\mathrm{e}^{-10}$.

(Note that $P(Y > 10) = 1 - F(10) \ne (1 - e^{-10})$).