# STA2001 Probability and Statistics (I)

Lecture 9

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

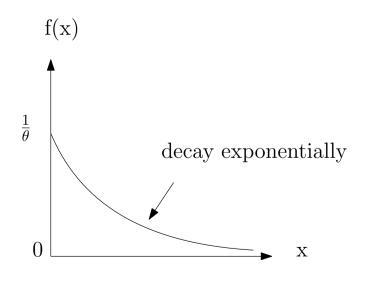
#### Review

Exponential distribution with parameter  $\theta = \frac{1}{\lambda}$ : X, the waiting time until the first occurrence in an approximate Poisson process with parameter  $\lambda > 0$  and its pdf takes the form of

$$first occur$$

$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \quad x \ge 0, \theta > 0$$

Mean and Variance:



$$E[X] = \theta, Var[X] = \theta^2$$

Mgf:

$$M(t) = \frac{1}{1-t\theta}, \quad t < \frac{1}{\theta}$$

#### Review

#### Definition

X, the waiting time until the  $\alpha$ th occurrence, and its pdf takes the form of

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\theta}}, \quad x \ge 0,$$

where  $\theta > 0$  and  $\alpha > 0$  are the two parameters,

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy, \quad t > 0$$

$$\Gamma(t) = (t-1)\Gamma(t-1), \qquad \Gamma(n) = (n-1)!$$

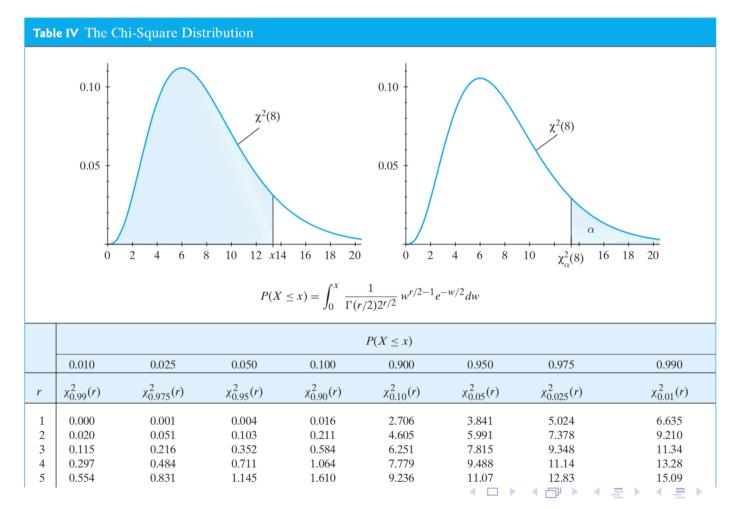
- $\alpha = 1$ , exponential distribution.  $\alpha = 1$  exponential distribution.
- $\theta = 2$ ,  $\alpha = \frac{r}{2}$ , r is an integer, chi square distribution (r is called the degrees of freedom). A = 2, A = 2

#### Review

Be able to calculate probabilities of events by looking up tables.

The tables of cdf of chi-square distribution are given

$$F(x) = P(X \le x) = \int_0^x f(t)dt.$$

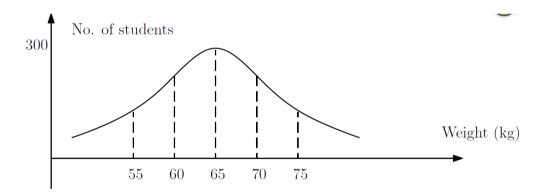


### 3.3 Normal Distribution

## **Description**

When observed over a large population, many things of interests have a "bell-shaped" relative frequency distribution.

- Weight of male students in CUHKsz
- Height
- TOFEL,IELTS test score



#### **Normal Distribution**

#### Definition

A continuous RV X is said to be normal or Gaussian if it has

a pdf of the form 
$$\int_{(x)} = \frac{1}{\sqrt{2\pi}6^2} e^{\frac{-(x-y)^2}{26^2}}$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}), \quad -\infty < x < \infty$$

where  $\mu$  and  $\sigma^2$  are two parameters characterizing the normal distribution. Briefly,  $X \sim N(\mu, \sigma^2)$ 

## pdf of Normal Distribution

f(x) is a well-defined pdf

1. 
$$f(x) > 0$$
 for all  $x$ .  $f(x) > 0$  for all  $x$ .

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

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

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

We will prove  $\int_{-\infty}^{\infty} f(x)dx = 1$  shortly, if time permits.  $\frac{dx}{dx} = \frac{dx}{6}$ 

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}6^{2}} \exp\left(-\frac{1}{2} \frac{(x-u)^{2}}{6^{2}}\right) dx = 1$$

Assume 
$$X \sim N(\mu, \sigma^2)$$
 
$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\lambda^2}{2}} d\lambda \int_{-\infty}^{\infty} e^{-\frac{\lambda^2}{2}} d\lambda$$

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$$

$$e^{tx} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} = \exp\left\{-\frac{1}{2\sigma^2}[x^2 - 2(\mu + \sigma^2 t)x + \mu^2]\right\} \left[\chi - (\mu + \delta^2 t)\right]^2$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{\lambda^2}{2}t\gamma^2} dy d\lambda$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{\lambda^2}{2}t\gamma^2} dy d\lambda$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{\lambda^2}{2}t\gamma^2} dy d\lambda$$

Consider

$$x^{2} - 2(\mu + \sigma^{2}t)x + \mu^{2} = [x - (\mu + \sigma^{2}t)]^{2} - 2\mu\sigma^{2}t - \sigma^{4}t^{2}$$

$$M(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \left[x - (\mu + \sigma^2 t)\right]^2\right) dx$$
$$\cdot \exp\left(\frac{-2\mu\sigma^2 t - \sigma^4 t^2}{-2\sigma^2}\right)$$

Recall that

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1$$
, independent of  $\mu$ 

Therefore,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} [x - (\mu + \sigma t)]^2\right) dx = 1$$

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \implies M(0) = 1;$$

$$M'(t) = (\mu + \sigma^2 t) \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \implies M'(0) = \mu$$

$$M''(t) = \sigma^2 \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) + (\mu + \sigma^2 t)^2 \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

$$\implies M''(0) = \mu^2 + \sigma^2$$

Recall that

$$E[X] = M'(0) = \mu$$

$$Var[X] = E[X^2] - (E[X])^2$$

$$=M''(0)-M'(0)^2=\sigma^2$$

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

$$E[X] = \mu, \quad Var[X] = \sigma^2$$

The two parameters  $\mu, \sigma^2$  are the mean and variance, respectively.

A RV X has its pdf in the form of

$$f(x) = \frac{1}{\sqrt{32\pi}} \exp\left[-\frac{(x+7)^2}{32}\right], -\infty < x < \infty$$

A RV X has its pdf in the form of

$$f(x) = \frac{1}{\sqrt{32\pi}} \exp\left[-\frac{(x+7)^2}{32}\right], -\infty < x < \infty$$

$$\Leftrightarrow$$
  $X \sim N(-7, 16)$ 

$$\Leftrightarrow X \sim N(-7, 16)$$

$$X \sim N(-7, 16)$$

$$U = -7 \quad G^2 = 16$$

$$\Leftrightarrow E(X) = -7, Var(X) = 16$$

$$\Leftrightarrow M(t) = \exp(-7t + 8t^2).$$

$$M(t) = \exp(ut + \frac{6^2}{2}t^2)$$

#### **Standard Normal Distribution**

Y is said to be a standard normal distribution if

$$Y \sim N(0,1) \Leftrightarrow \text{ its pdf } f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

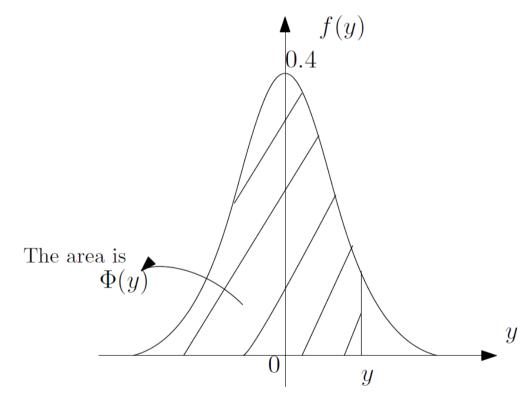
$$Y \sim N(0,1) \Leftrightarrow \text{pdf } f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$
Its cdf

$$\Phi(y) = P(Y \le y) = \int_{-\infty}^{y} f(z) dz = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Due to the symmetry of f(y),  $\Phi(-y) = 1 - \Phi(y)$ , for any y

# pdf of N(0,1)

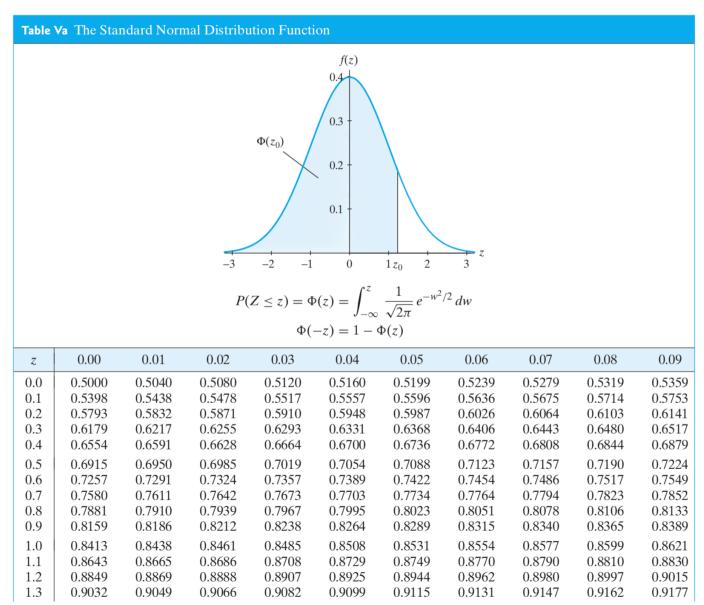
$$Y \sim N(0,1) \Leftrightarrow ext{ its pdf } f(y) = rac{1}{\sqrt{2\pi}} e^{-rac{y^2}{2}}$$



Due to the symmetry of f(y),  $\Phi(-y) = 1 - \Phi(y)$ , for any y

# pdf of N(0,1)

Values of  $\Phi(y)$  for values of  $y \ge 0$  are in Appendix B (page 502).



# pdf of N(0,1)

Values of  $\Phi(y)$  for values of  $y \ge 0$  are in Appendix B (page 502).

1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

 $Z \sim N(0,1)$  Then compute

$$P(Z \le 1.24) = \Phi(1.24) = 0.8925$$

$$P(1.24 \le Z \le 2.37) = \Phi(2.37) - \Phi(1.24)$$

 $Z \sim N(0,1)$  Then compute

$$P(Z \le 1.24) = \Phi(1.24) = 0.8925$$

$$P(1.24 \le Z \le 2.37) = \Phi(2.37) - \Phi(1.24)$$

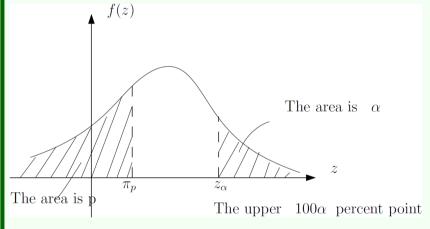
$$= 0.9911 - 0.8925 = 0.0986$$

$$P(-2.37 \le Z \le -1.24) = P(1.24 \le Z \le 2.37) = 0.0986$$

# The upper $100\alpha$ percent point

#### Definition

The number  $z_{\alpha}$  such that  $P(Z \geq z_{\alpha}) = \alpha$ .



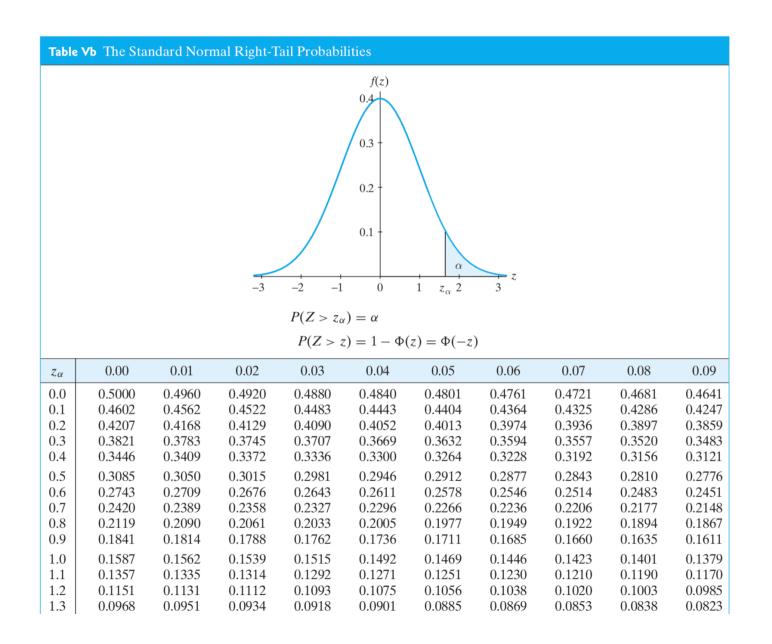
 $P(X \le \pi_p) = p$ ,  $\pi_p$  is 100pth percentile.

Note:

$$P(Z < z_{\alpha}) = 1 - P(Z \ge z_{\alpha})$$
  
= 1 - \alpha

So  $z_{\alpha}$  is the  $100(1-\alpha)$ th percentile

## The upper $100\alpha$ percent point



 $Z \sim N(0,1)$ , Find  $Z_{0.0125}$ . That is

$$P(Z \ge z_{0.0125}) = 0.0125$$

 $Z \sim N(0,1)$ , Find  $Z_{0.0125}$ . That is

$$P(Z \ge z_{0.0125}) = 0.0125$$

check the table  $\Rightarrow z_{0.0125} = 2.24$ 

What about  $z_{0.05}$  and  $z_{0.025}$ ?

 $Z \sim N(0,1)$ , Find  $Z_{0.0125}$ . That is

$$P(Z \ge z_{0.0125}) = 0.0125$$

check the table  $\Rightarrow z_{0.0125} = 2.24$ 

What about  $z_{0.05}$  and  $z_{0.025}$ ?

$$\Rightarrow z_{0.05} = 1.645, \quad z_{0.025} = 1.960$$

Now we know how to compute  $\Phi(y)$  by looking up the table for  $Y \sim N(0,1)$ .

 $Z \sim N(0,1)$ , Find  $Z_{0.0125}$ . That is

$$P(Z \ge z_{0.0125}) = 0.0125$$

check the table  $\Rightarrow z_{0.0125} = 2.24$ 

What about  $z_{0.05}$  and  $z_{0.025}$ ?

$$\Rightarrow z_{0.05} = 1.645, \quad z_{0.025} = 1.960$$

Now we know how to compute  $\Phi(y)$  by looking up the table for  $Y \sim N(0,1)$ . What if Y is not standard normal?

#### Theorem 3.3-1

#### Theorem

If Y is  $N(\mu, \sigma^2)$ , then  $X = \frac{Y - \mu}{\sigma}$  is N(0, 1)

Proof: The idea is to show X has the same cdf as N(0,1)

$$P(X \le x) = P(\frac{Y - \mu}{\sigma} \le x) = P(Y \le \sigma x + \mu) = \int_{-\infty}^{\sigma x + \mu} f(y) dy$$
$$= \int_{-\infty}^{\sigma x + \mu} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(y - \mu)^2}{\sigma^2}\right) dy$$

#### Theorem 3.3-1

coordinate change

$$w = \frac{y - \mu}{\sigma} \Longrightarrow \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}w^2) dw. \longrightarrow \text{cdf of } N(0, 1).$$

Therefore,  $P(X \le x) = \Phi(x)$  and this completes the proof.

With the above theorem, for  $X \sim N(\mu, \sigma^2)$ 

$$P(a \le X \le b) = P(\frac{a - \mu}{\sigma} \le \frac{X - \mu}{\sigma} \le \frac{b - \mu}{\sigma})$$
$$= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

where  $\Phi(\cdot)$  is the cdf of N(0,1).

 $X \sim N(3, 16)$  Compute  $P(4 \le X \le 8)$ ,  $P(0 \le X \le 5)$ .

$$X \sim N(3, 16)$$
 Compute  $P(4 \le X \le 8)$ ,  $P(0 \le X \le 5)$ .

$$P(4 \le X \le 8) = P(\frac{4-3}{4} \le \frac{X-3}{4} \le \frac{8-3}{4})$$

$$= \Phi(1.25) - \Phi(0.25) = 0.8944 - 0.5987$$

$$P(0 \le X \le 5) = P(\frac{0-3}{4} \le \frac{X-3}{4} \le \frac{5-3}{4})$$

$$= \Phi(0.5) - \Phi(-0.75) = 0.6915 - 0.2266.$$

# Relation between normal and $\chi^2$ distribution

#### Theorem 3.3-2

If X is  $N(\mu, \sigma^2)$  with  $\sigma^2 > 0$ , then

$$\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$$

Proof: Let  $V = \frac{(X-\mu)^2}{\sigma^2}$ . Then consider the cdf of V,

$$G(v) = P(V \le v) = P(-\sqrt{v} \le Z \le \sqrt{v})$$

where 
$$Z = \frac{X - \mu}{\sigma}$$
, with  $\nu \geq 0$ .

$$\sqrt{=\frac{(X-N)^2}{(X-N)^2}}$$

th 
$$v \ge 0$$
.

$$V = \frac{(\chi - u)^2}{6}$$

$$Z = \frac{\chi - u}{6}$$

$$V = Z^2$$

# Relation between normal and $\chi^2$ distribution

$$G(v) = \int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 2 \int_{0}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Change the variable of integration  $z = \sqrt{y}$   $\frac{dz}{dy} = \frac{1}{2\sqrt{y}}$ 

$$z = \sqrt{y}$$
  $\frac{dz}{dy} = \frac{1}{2\sqrt{y}}$ 

$$G(v) = 2 \int_0^v \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} dy = \int_0^v \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} dy, v \ge 0$$

$$g(v) = G'(v) = \frac{1}{\sqrt{2\pi}} v^{\frac{1}{2} - 1} e^{-\frac{1}{2}v}, v \ge 0. \quad \int_{0}^{\sqrt{1}} \int_{2\pi}^{\sqrt{2}} e^{-\frac{1}{2}z^{2}} dz$$
Now recall the pdf of  $\chi^{2}(1)$ :
$$Z = \int_{0}^{\sqrt{2}} \int_{2\pi}^{\sqrt{2}} dz$$

$$Z = \int_{0}^{\sqrt{2}} \int_{2\pi}^{\sqrt{2}} dz$$

$$T(\frac{1}{2}) 2^{\frac{1}{2}} x^{-\frac{1}{2}} e^{-\frac{1}{2}x}, \quad x \ge 0,$$

$$f(x) = \frac{1}{\Gamma(\frac{1}{2})2^{\frac{1}{2}}} x^{-\frac{1}{2}} e^{-\frac{1}{2}x}, \quad x \ge 0,$$

#### Theorem 3.3-2

Since g(v) is a pdf, then  $\int_0^\infty g(v)dv = 1$ 

$$1 = \int_0^\infty \frac{1}{\sqrt{2\pi}} v^{\frac{1}{2} - 1} e^{-\frac{1}{2}v} dv \stackrel{x = \frac{1}{2}v}{=} \frac{1}{\sqrt{\pi}} \int_0^\infty x^{\frac{1}{2} - 1} e^{-x} dx$$

$$= \frac{1}{\sqrt{\pi}} \Gamma(\frac{1}{2}) \Rightarrow \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

$$\implies g(v) = \frac{1}{\Gamma(\frac{1}{2})2^{\frac{1}{2}}} v^{\frac{1}{2}-1} e^{-\frac{v}{2}}, v > 0$$