ISyE 3770, Spring 2024
Statistics and Applications

Introduction to Continuous Distribution (II)

Instructor: Jie Wang
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Tech

jwang3163@gatech.edu
Office: ISyE Main 445

Section 3.2 exponential, gamma, chi-square Distributions

Definition [chi-square distribution]

Let X have a Gamma distribution with $\theta = 2, \alpha = \frac{r}{2}$, where r is an integer. The pdf of X is $f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, x > 0.$

Then X has **chi-square distribution** with $\boldsymbol{r}$ degrees of freedom, which is denoted by $X \sim \chi^2(r)$.

➢Mean and Variance

Substitute the α and θ in mean and variance of **Gamma distribution** into it.

$$\mathbb{E}[X] = \alpha\theta = r, \qquad \mathrm{Var}(X) = \alpha\theta^2 = 2r,$$

$$\text{mgf: } M(t) = \left(\frac{1}{1-\theta t}\right)^{\alpha} = (1-2t)^{-r/2}, \qquad t < \frac{1}{2}.$$

Remark: chi-square distribution plays an important role in Statistics. The tables of the values for *cdf* of chi-square distribution are given in our textbook!

$$F(x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2} dw.$$

for selected values of r and x. (Please check *Table IV in Appendix A* on textbook)

Example 2

Let X have a chi-square distribution with r=5 degrees of freedom. Then find $P(1.145 \leq X \leq 12.83)$ and $P(X > 15.09)$.

$$\chi^2_{0.025}(5) = 12.83,$$
$$\chi^2_{0.95}(5) = 1.145,$$
$$\chi^2_{0.01}(5) = 15.09$$

Solution:

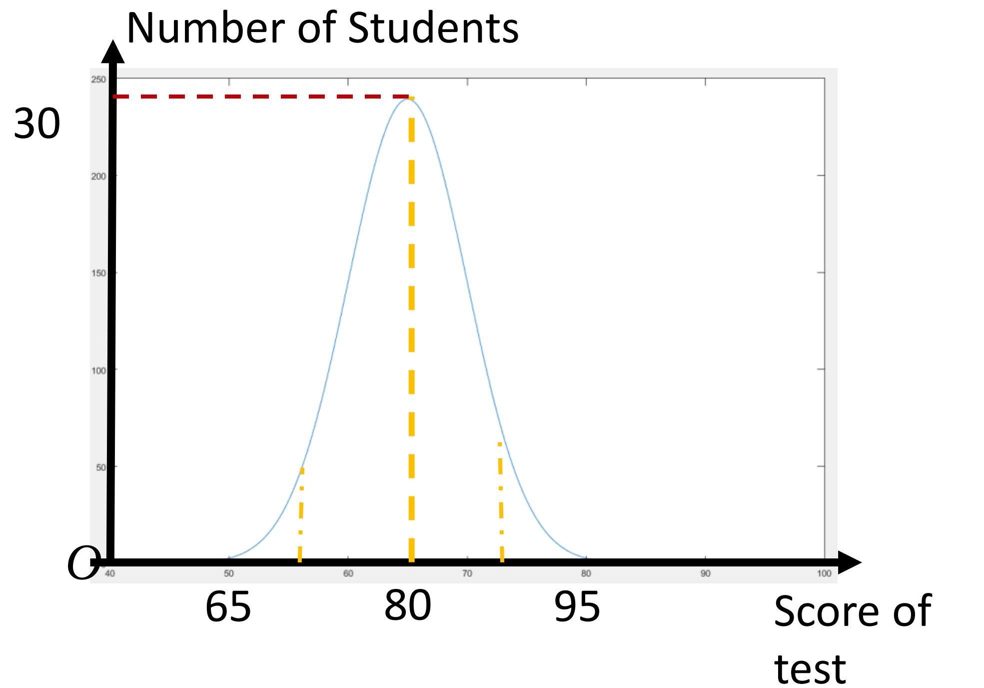
$$\begin{aligned} P(1.145 \leq X \leq 12.83) =& F(12.83) - F(1.145) \\ =& (1 - 0.025) - (1 - 0.95) = 0.925 \\ P(X > 15.09) =& 1 - F(15.09) = 1 - (1 - 0.01) = 0.01 \end{aligned}$$

Chapter 3 Continuous distribution

Section 3.3 Normal distribution

Situation: When observed over a large population, many variables have a “bell-shaped” relative frequency distribution.

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



A very useful family of probability distributions for such variables are the normal distributions.

Definition [Normal distribution]

A continuous *RV X* is said to be ***normal*** or ***Gaussian*** if has a pdf of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right), \qquad x \in (-\infty, +\infty),$$

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

What's interpretation of μ and σ^2 ? consider mean and Variance.

➢ *f*(*x*) is a well-defined *pdf*

- $f(x) \geq 0$ for all x.
- We need to check whether $\int_{-\infty}^{\infty} f(x)dx = 1$.

  We take $I = \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx$.

  By change of variable, we take $z = \frac{x-\mu}{\sigma}$. Then

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz.$$

Since $I > 0$, it suffices to check $I^2 = 1$.

➢ *f*(*x*) is a well-defined *pdf* (c.n.t.)

$$I^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-z^2/2}dz \int_{-\infty}^{+\infty} e^{-y^2/2}dy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-z^2/2}e^{-y^2/2}dzdy = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(y^2+z^2)/2}dydz$$

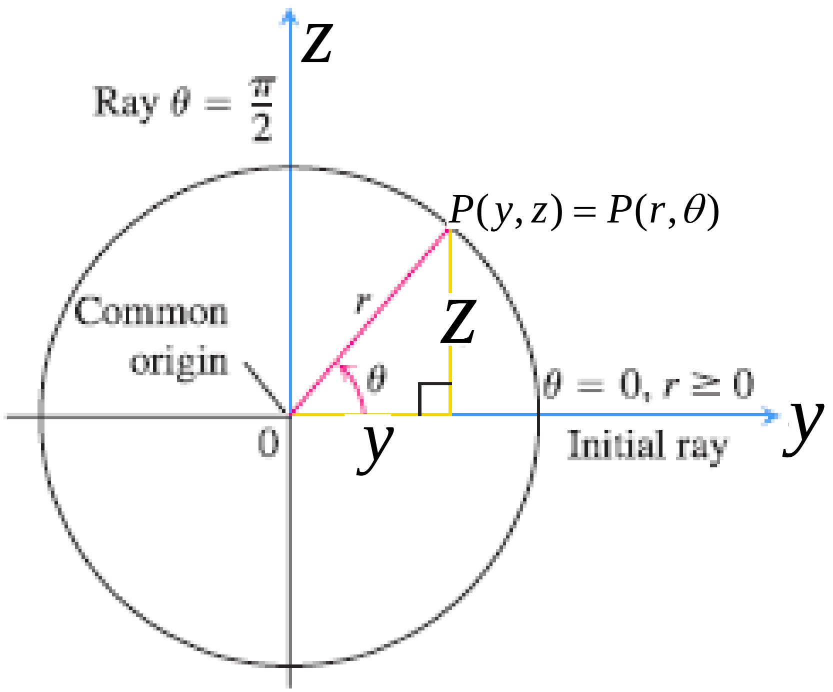
$$\text{Coordinate change}: \begin{cases} y = r\cos\theta \\ z = r\sin\theta \end{cases}$$

(polar coordinate)

$$I^2 = \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{+\infty} e^{-r^2/2}rdrd\theta$$

$$= \frac{1}{2\pi}\int_{0}^{2\pi} d\theta \cdot \int_{0}^{+\infty} e^{-r^2/2}rdr = 1$$

Thus, $I = 1$, and we have shown that $f(x)$ has the properties of a pdf.



If you don't know some specific steps to derive this conclusion, memory is a good solution.

➢ Mean and Variance (mgf approach)

Assume $X \sim N(\mu, \sigma^2)$, then $M(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} dx$

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

Note that $x^2 - 2(\mu + \sigma^2 t)x + \mu^2 = \left[x - (\mu + \sigma^2 t)\right]^2 - 2\mu\sigma^2 t - \sigma^4 t^2$.

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

Recall for any μ, it holds that $I = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1$.

Substituting μ with $\mu + \sigma^2 t$ implies

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

$$M(t) = \exp\left(\frac{-2\mu\sigma^2 t - \sigma^4 t^2}{-2\sigma^2}\right) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

How to derive the mean and Variance based on *mgf* ?

➢ Mean and Variance (c.n.t.)

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

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

⇒

$$M'(0) = \mu = \mathbb{E}[X]$$
$$M''(0) = \mu^2 + \sigma^2 = \mathbb{E}[X^2]$$
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sigma^2$$

Remark: For $X \sim N(\mu, \sigma^2)$, it holds that $E[X] = \mu$, $Var(X) = \sigma^2$.

Example 1

A RV X has the pdf

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

Calculate the mgf of X.

Solution. One can check that $X \sim N(-7, 16)$, then

$$\mathbb{E}[X] = -7, \mathrm{Var}(X) = 16.$$

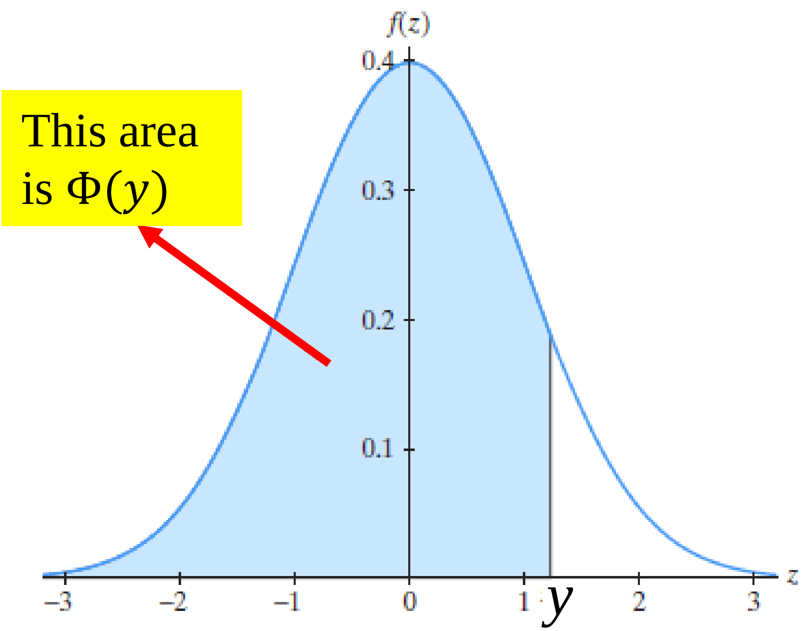
Hence, we obtain its mgf $M(t) = \exp(-7t + 8t^2)$.

Definition [Standard normal distribution]

A RV Y is said to have a **standard normal distribution** if $Y \sim N(0,1)$, i.e., its pdf is

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

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



- Values of $\Phi(y)$ for some values of $\boldsymbol{y \ge 0}$ are given in Appendix A in our textbook!
- Due to the ***symmetry*** of $f(y)$,

$$\Phi(y) = 1 - \Phi(-y)$$

for all real y.

Example 2

Take $Z \sim N(0,1)$, then compute

$$P(Z \le 1.24), P(1.24 \le Z \le 2.37), P(-2.37 \le Z \le -1.24),$$
$$P(Z > 1.24), P(Z \le -2.14), P(-2.14 \le Z \le 0.77)$$

Solution:

Using Table provided in Appendix, we have:

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

$P(1.24 \le Z \le 2.37) = \Phi(2.37) - \Phi(1.24) = 0.9911 - 0.8925 = 0.0986$

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

Using Table in Appendix, we have:

$P(Z > 1.24) = 0.1075$

$P(Z \le -2.14) = P(Z \ge 2.14) = 0.0162$

Finally, we have:

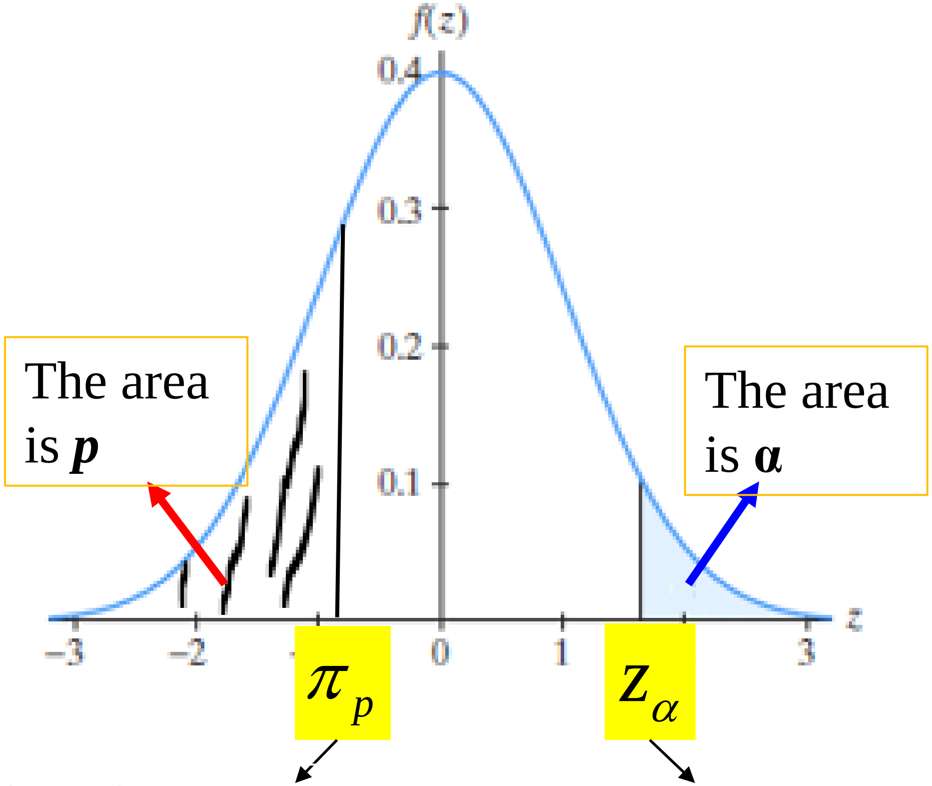
$P(-2.14 \le Z \le 0.77) = P(Z \le 0.77) - P(Z \le -2.14) = 0.7794 - 0.0162 = 0.7632.$

given a probability p, we can also find a constant a so that $P(Z \le a) = p$ through using the table!.

Definition [the upper 100α percent point]

It is a number z_α such that the area **under $f(x)$ to the right of z_α** is $\boldsymbol{\alpha}$. That is,

$$P(Z \ge z_\alpha) = \alpha$$



Note that

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

So z_α is the **(100(1-α))th percentile**.

(100p)th percentile

the upper 100α percent point

$\mathrm{P}(X \le \pi_p) = p, \pi_p$ is (100p)th percentile.

Example 3

$Z \sim N(0,1)$, Find $z_{0.0125}$、 $z_{0.05}$、 $z_{0.025}$.

Solution:

$\Leftrightarrow P(Z \geq z_{0.0125}) = 0.0125.$ By checking the Table in book, $z_{0.0125} = 2.24.$

Similarly, $z_{0.05} = 1.645$, $z_{0.025} = 1.960.$

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

Theorem If Y is $N(\mu, \sigma^2)$, then $X = (Y-\mu)/\sigma$ is $N(0,1)$.

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

$$P(X \leq x) = P\left(\frac{Y-\mu}{\sigma} \leq x\right) = P(Y \leq \sigma x + \mu) = \int_{-\infty}^{\sigma x + \mu} f(y)\,dy$$

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

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} w^2\right) dw = \Phi(x)$$

Change of variable with $w = \dfrac{y-\mu}{\sigma}$

cdf of N(0,1)

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

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

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

Example 4

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

Solution:

$$P(4 \le X \le 8) = P\left(\frac{4-3}{4} \le \frac{X-3}{4} \le \frac{8-3}{4}\right) = \Phi(1.25) - \Phi(0.25) = 0.8944 - 0.5987 = 0.2957.$$

$$P(0 \le X \le 5) = P\left(\frac{0-3}{4} \le \frac{X-3}{4} \le \frac{5-3}{4}\right) = \Phi(0.5) - \Phi(-0.75) = 0.6915 - 0.2266 = 0.4649.$$

In the next theorem, we give a relationship between the chi-square and normal distributions.

Theorem

If the RV X is $N(\mu, \sigma^2)$ with $\sigma^2 > 0$, then $\dfrac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$.

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

$$G(v) = P(V \le v) = P(-\sqrt{v} \le Z \le \sqrt{v}) \text{ with } Z = \frac{X-\mu}{\sigma}, v \ge 0.$$

$$G(v) = \int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 2\int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

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

Changing of variable with $z = \sqrt{y}$ and $\dfrac{dz}{dy} = \dfrac{1}{2\sqrt{y}}$:

The pdf of V is: $g(v) = G'(v) = \dfrac{1}{\sqrt{2\pi v}} e^{-\frac{1}{2}v}, v \ge 0.$ since $g(v)$ is a pdf, $\int_0^{\infty} g(v)dv = 1.$

$$\Rightarrow 1 = \int_0^{\infty} \frac{1}{\sqrt{2\pi v}} e^{-\frac{1}{2}v} dv = \frac{1}{\sqrt{\pi}} \int_0^{\infty} \frac{1}{\sqrt{x}} e^{-x} dx = \frac{1}{\sqrt{\pi}} \int_0^{\infty} x^{1/2-1} e^{-x} dx = \frac{1}{\sqrt{\pi}} \Gamma(\frac{1}{2})$$

$$\Rightarrow \Gamma(\frac{1}{2}) = \sqrt{\pi} \Rightarrow g(v) = \frac{1}{\Gamma(\frac{1}{2})2^{1/2}} v^{1/2-1} e^{-\frac{1}{2}v}, \quad v > 0.$$

$$\Rightarrow V \sim \chi^2(1)$$