Math 4501 - Probability and Statistics II

new content

3.2 - The chi-square distribution
(as an element of the gamma distribution family)

review

Chi-square distribution

$\alpha = \frac{r}{2}$

Let X have a gamma distribution with parameters $\theta = 2$ and $\alpha = r/2$, where r is a positive integer.

The pdf of X is

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} \mathrm{e}^{-x/2}, \quad x > 0$$

and we say that X has a chi-square distribution with r degrees of freedom.

gamma with
$\theta = 2$
$\alpha = \frac{r}{2}$, $r \in \mathbb{N}$
↑
d.o.f

Notation and terminology: We say that X is $\chi^2(r)$ and denote it as $\boxed{X \sim \chi^2(r).}$

X follows a χ^2 distribution with r d.o.f.

Notes:

- $\chi^2(r)$ is a gamma distribution with $\theta = 2$ and $\alpha = r/2$, where $r \in \mathbb{N}$.
- $\chi^2(2)$ is an exponential distribution with $\theta = 2$.

if $r = 2 \Rightarrow \alpha = \frac{2}{2} = 1 \Rightarrow \chi^2(2)$ is $\mathrm{Exp}(\theta = 2)$

Moment generating function for chi-square distribution

Suppose X as a chi-square distribution with r degrees of freedom.

The mgf of X is given by

$$M(t) = \frac{1}{(1-2t)^{r/2}}, \qquad t < \frac{1}{2}$$

The mean and variance of X are

- $\mu = E[X] = \frac{r}{2}2 = r$
- $\sigma^2 = \text{Var}(X) = \frac{r}{2}2^2 = 2r$

Percentiles for chi-square distribution

Suppose X is $\chi^2(r)$, and let $\alpha \in (0, 1)$ (usually $\alpha < 0.5$).

The $100(1-\alpha)$th percentile is the number $\chi^2_\alpha(r)$ such that

$$P\left[X \geq \chi^2_\alpha(r)\right] = \alpha \ ,$$

that is, the probability to the right of $\chi^2_\alpha(r)$ is α.

The 100α percentile is the number $\chi^2_{1-\alpha}(r)$ such that

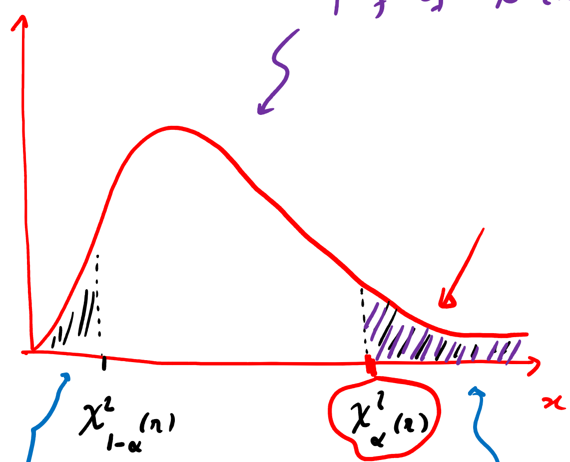
$$P\left[X \leq \chi^2_{1-\alpha}(r)\right] = \alpha \ ,$$

that is, the probability to the right of $\chi^2_{1-\alpha}(r)$ is $1-\alpha$.

$\chi^2_\alpha(r)$

same shape as gamma distn. pdf

pdf of $\chi^2(r)$



shaded area to the left is α

shaded area to the right is α

Example

Suppose X is $\chi^2(7)$. Find two constants, a and b, such that

$$\boxed{P(a < X < b) = 0.95}$$

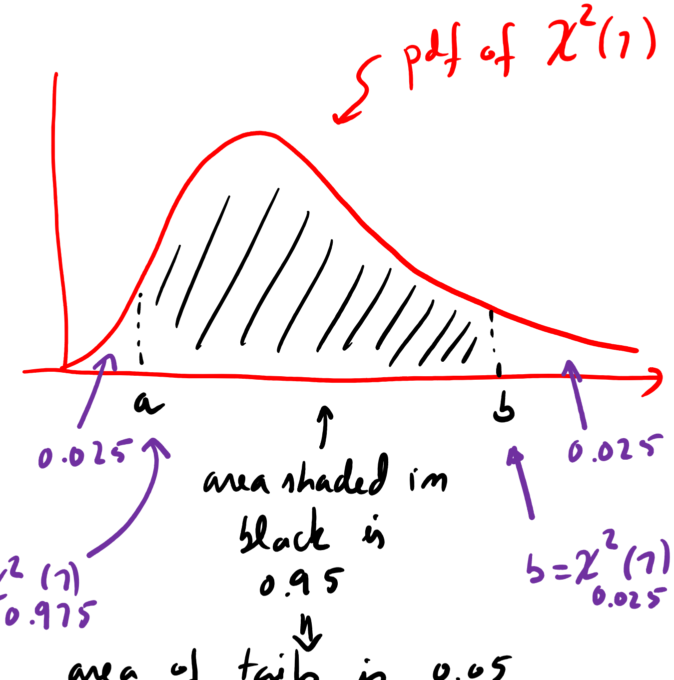
important example / technique for confidence intervals (to be studied later)

$X \sim \chi^2(7)$ ← chi-squared distri. with $n = 7$ d.o.f.

Note that there are ∞-many possible choices for a and b depending on how much area we want to assign to the left and right tails (the only constraint being that the sum of the area of the two tails is 0.05)

We may pick the most "symmetric" choice for a and b, that is, make the left and right tails have both area

$$0.025 \leftarrow \frac{1 - 0.95}{2} \Rightarrow \begin{cases} a = \chi^2_{0.975}(7) = 1.69 \\ b = \chi^2_{0.025}(7) = 16.01 \end{cases}$$



Example

Suppose customers arrive at a shop on the average of 30 per hour in accordance with a Poisson process. Determine the probability that the shopkeeper will have to wait longer than 9.390 minutes for the first nine customers to arrive.

Let X be the r.v. representing the time (in minutes) for the first nine customers to arrive.

$\Downarrow$

X has a gamma distribution with parameters $\overbrace{\alpha = 9}^{\text{wait-time for } 9^{th} \text{ client}}$ and

average number of clients per unit time (from the Poisson process!) $\rightarrow \lambda = \dfrac{30 \text{ customers}}{1 \text{ hour}} = \dfrac{30 \text{ customers}}{60 \text{ minutes}} = \dfrac{1}{2}$ customer/minute

Thus, the mean wait time between consecutive arrivals is $\theta = \dfrac{1}{\lambda} = \dfrac{1}{1/2} = 2$ (minutes)

CONCLUSION: X has a gamma distribution with parameters $\alpha = 9$ and $\theta = 2$

Since $\theta = 2$ and $\alpha = 9$ is of the form $\frac{r}{2}$ with $r = 18$

$$\underbrace{\frac{r}{2}}_{\alpha} = 9$$

then this distribution is also a $\chi^2(\underset{\uparrow r}{18})$

Question is then to find the probability:

$$P(X > 9.39) = 1 - P(X \leq 9.39) = 1 - 0.05 = 0.95$$

table for $\chi^2(18)$

Math 4501 - Probability and Statistics II

3.3 - The chi-square distribution
(connection with the normal distribution)

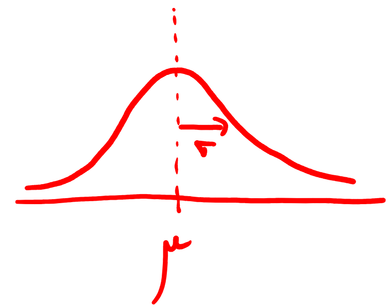
Normal distribution (review)

mean

standard deviat.

A random variable X has a normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ if its pdf is of the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] , \quad x \in \mathbb{R} .$$



Notation: We say that X is $N\left(\mu, \sigma^2\right)$ and denote it as $X \sim N\left(\mu, \sigma^2\right)$

Note: It is possible to check that:

- $f(x) > 0$ for all $x \in \mathbb{R}$
- $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$

Normal distribution mgf, mean and variance (review)

Suppose X has a normal distribution (with parameter μ and σ).

The mgf of X is given by

$$M(t) = E\left[\mathrm{e}^{tX}\right] = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) , \quad x \in \mathbb{R} .$$

The mean and variance are

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

Standard normal distribution (review)

$\mu = 0$

We say that Z has a standard normal distribution if Z is $N(0, 1)$.

$\sigma^2 = 1$

The pdf of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}$$

and its cdf is

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-w^2/2}\, \mathrm{d}w$$

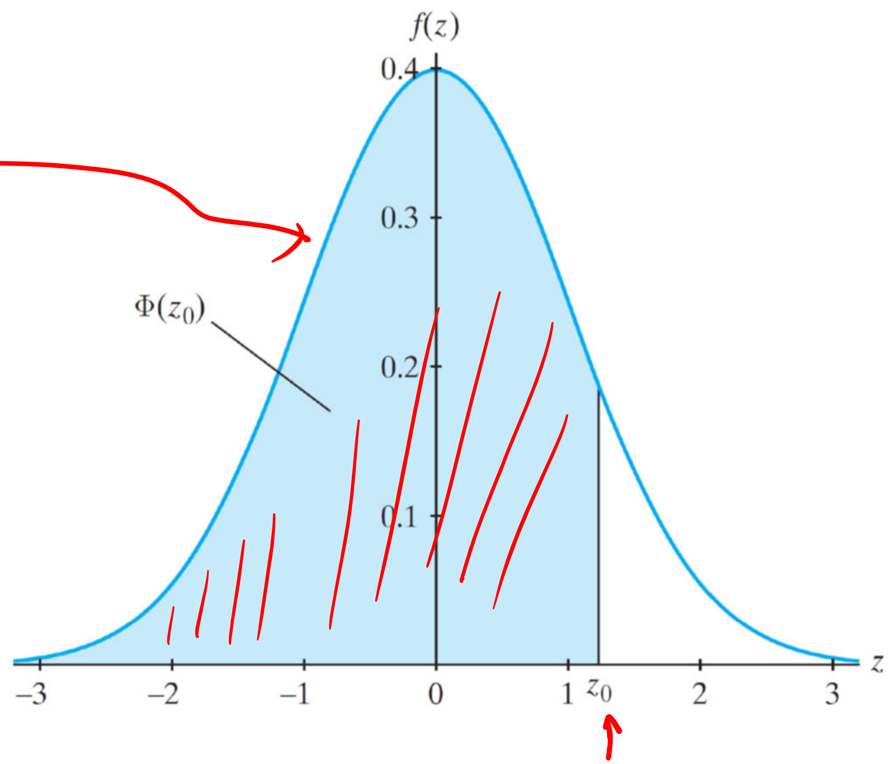
Notes:

- It is not possible to evaluate this integral by finding an antiderivative that can be expressed as an elementary function.
- Numerical approximations for integrals of this type have been tabulated. (Tables Va and Vb in textbook Appendix B)

The graph of the pdf of Z is a bell-shaped curved.

$Z \sim N(0,1)$

$\Phi(z_0)$

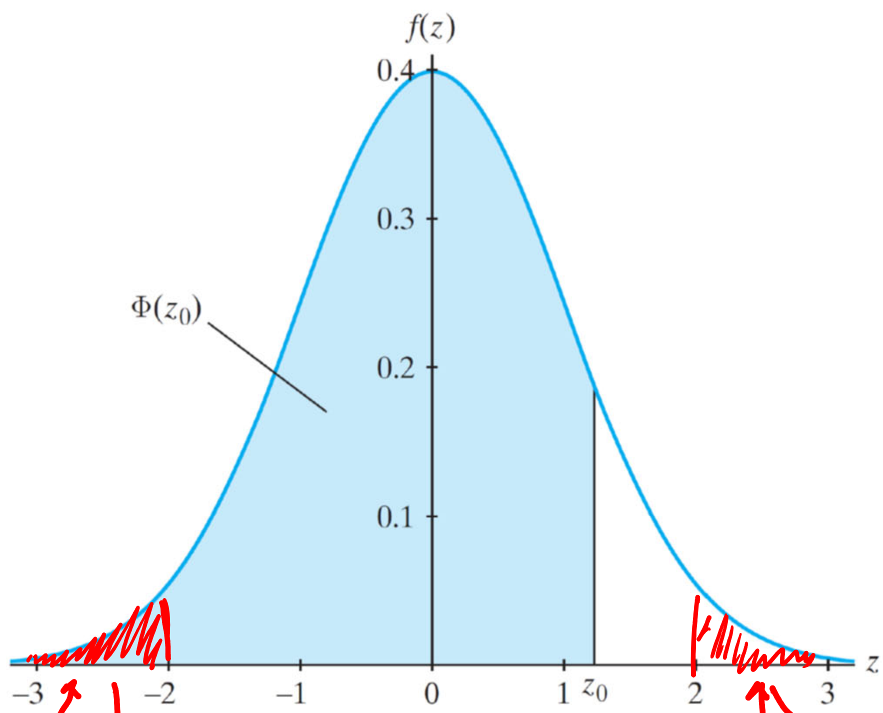


The shaded area equals $\Phi(z_0) = P(Z \leq z_0)$.

Symmetry guarantees that for all real z, we have

$$\Phi(-z) = P(Z \le -z) = P(Z > z) = 1 - P(Z \le z) = 1 - \Phi(z) \ .$$

$\phi(-z) = 1 - \phi(z)$



Example

Suppose Z is $N(0,1)$. $Z \sim N(0,1)$ standard normal

Evaluate each of the following probabilities:

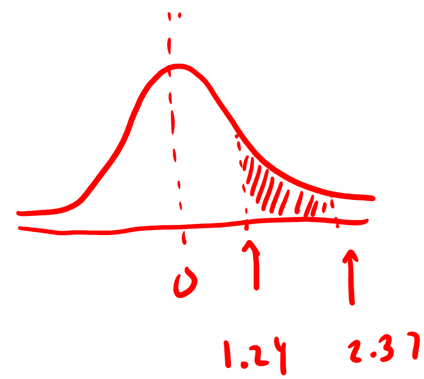
- $P(Z \leq 1.24)$
- $P(1.24 \leq Z \leq 2.37)$
- $P(-2.37 \leq Z \leq -1.24)$

cdf of Z is

$$\phi(z) = P(Z \leq z)$$

$$P(Z \leq 1.24) = \Phi(1.24) \overset{\text{table}}{=} 0.8925$$

$$P(1.24 < Z < 2.37) = \underbrace{\underbrace{\phi(2.37)}_{\text{area to the left of } 2.37} - \underbrace{\phi(1.24)}_{\text{area to the left of } 1.24}}_{\text{area between } 1.24 \text{ and } 2.37} = 0.9911 - 0.8925$$



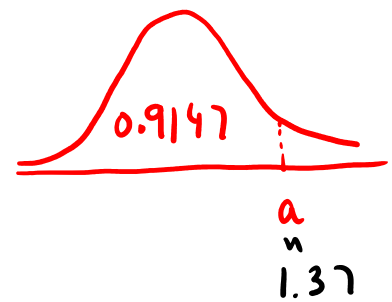
Example

Suppose Z is $N(0,1)$. $Z \sim N(0,1)$

Find constants a and b such that

$$\underbrace{P(Z \le a)}_{\text{area to the left of } a} = 0.9147 \quad \text{and} \quad \underbrace{P(Z \ge b)}_{\text{area to right of } b} = 0.0526 .$$

From the table we get $a = 1.37$

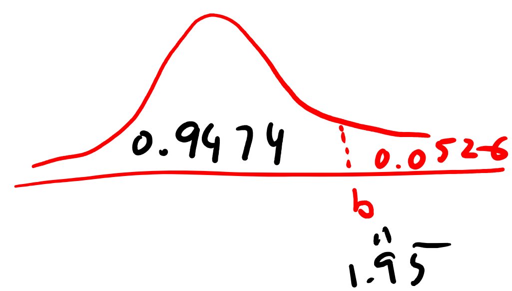


For the 2nd item:

$$P(Z \geq b) = 0.0526 \iff 1 - P(Z < b) = 0.0526$$

$$\iff P(Z < b) = 1 - 0.0526 \iff P(Z<b) = 0.9474$$

$$b = 1.95$$



Percentiles for standard normal distribution (review)

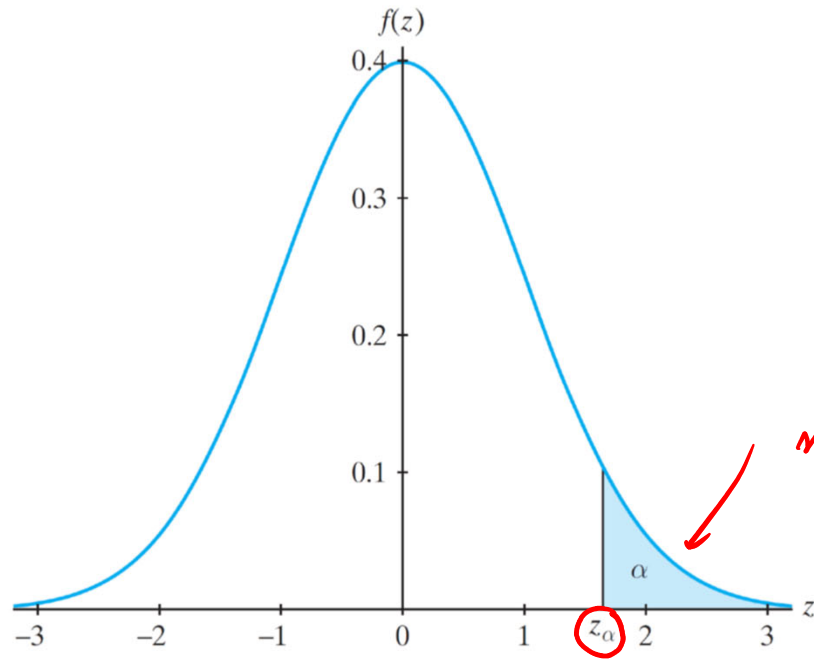
Suppose Z is $N(0,1)$, and let $\alpha \in (0,1)$ (usually $\alpha < 0.5$).

The $100(1-\alpha)$th percentile of its distribution is the number z_α such that

$$P[Z \geq z_\alpha] = \alpha \; ,$$

$\leftarrow$ tail to the right has area α

that is, the probability to the right of z_α is α.



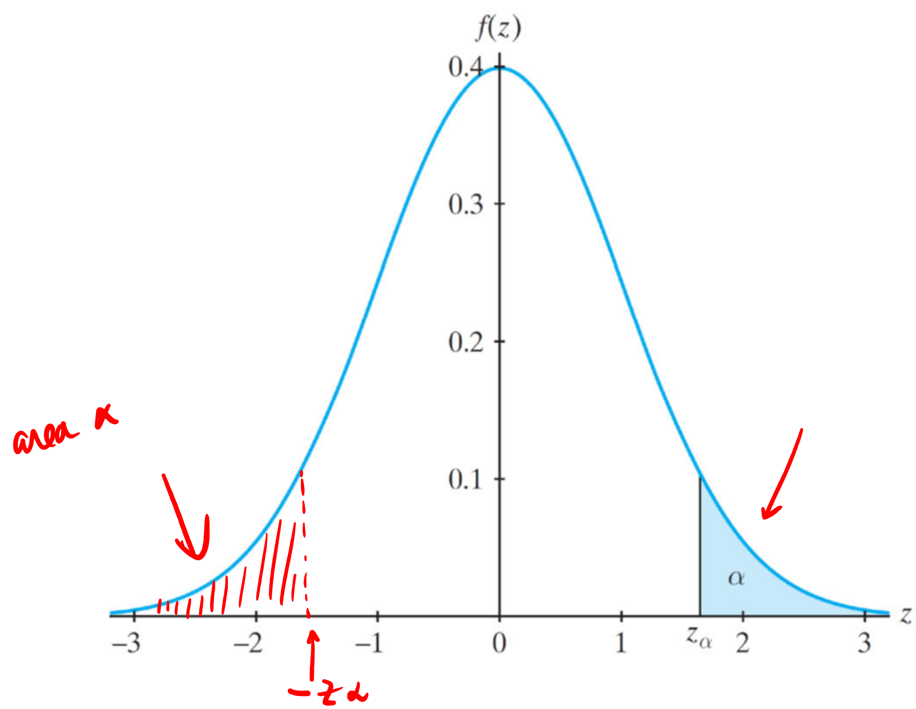
shaded region has area α

Due to the symmetry of the standard normal pdf, we have

$$P\left(Z \leq -z_\alpha\right) = P\left(Z \geq z_\alpha\right) = \alpha$$

and, since the subscript of z_α is the right-tail probability, we also have

$$z_{1-\alpha} = -z_\alpha \ .$$



Example

Find:

- $z_{0.0125}$
- $z_{0.05}$ $= 1.645 \leftarrow$ from table (row for z_α)

$z_{0.0125} = 2.240$

$\uparrow$

from the row for $z_{\frac{\alpha}{2}}$ at the bottom of the table

when $\alpha = 0.025$ so that $\frac{\alpha}{2} = 0.0125$

Standard score (review)

Theorem

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

$$\boxed{Z = \frac{X-\mu}{\sigma} \sim N(0,1)}$$

is $N(0,1)$.

Notes:

1. If X is normally distributed, then Z is normally distributed with zero mean and unit variance.

2. Z is often called the standard score associated with X.

3. Standard scores can be used to find probabilities associated with $X \sim N(\mu, \sigma^2)$:

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \overbrace{\frac{X-\mu}{\sigma}}^{Z} \leq \frac{b-\mu}{\sigma}\right) \overset{\text{table!}}{=} \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Example

Suppose X is $N(3, 16)$. $\Rightarrow \mu = 3$ and $\sigma^2 = 16 \Rightarrow \sigma = 4$

σ^2

Evaluate each of the following probabilities:

- $P(4 \leq X \leq 8)$
- $P(0 \leq X \leq 5)$

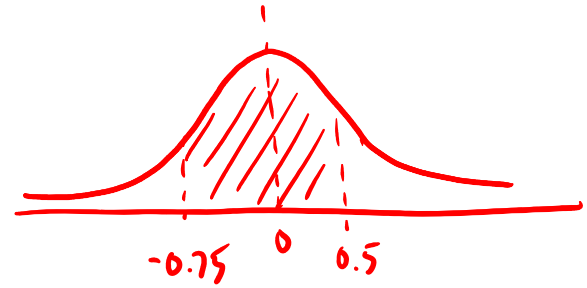
By the previous theorem, if $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim N(0,1)$

In this example, we have $Z = \dfrac{X-3}{4} \sim N(0,1)$

$$P(0 \leq X \leq 5) = P\left(\frac{0-3}{4} \leq \frac{X-3}{4} \leq \frac{5-3}{4}\right)$$

$$= P\left(-\frac{3}{4} \leq Z \leq \frac{5-3}{4}\right) = P(-0.75 \leq Z \leq 0.5) = ??$$

$$P(-0.75 \le Z \le 0.5) = \phi(0.5) - \underbrace{\phi(-0.75)}_{\text{does not show up on table!}}$$



use symmetry:

$$\phi(-0.75) = 1 - \phi(0.75)$$

$$= \phi(0.5) - (1 - \phi(0.75))$$

$$= \phi(0.5) + \phi(0.75) - 1$$

table

$$= 0.6915 + 0.7734 - 1$$

$$= \text{----}$$

Relationship between standard normal and chi-square

Theorem (VERY IMPORTANT!)

If X is $N(\mu, \sigma^2)$, then the random variable

$$V = Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$$

is $\chi^2(1)$.

$$X \sim N(\mu, \sigma^2) \xrightarrow[\text{previous thm}]{} Z = \frac{X-\mu}{\sigma} \sim N(0,1) \xrightarrow[\text{theorem above}]{} V = Z^2 \sim \chi^2(1)$$

<u>Proof</u> We know from the previous theorem (from 3501) that if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Thus, we only need to prove that

$$V = Z^2 \sim \chi^2(1) \quad \text{whenever} \quad Z \sim N(0,1)$$

Since $V = Z^2$, than V takes only non negative values:

so $P(V \leq 0) = 0$

Take $y > 0$ and let us determine the cdf (and later the pdf) of the r.v. $V = Z^2$

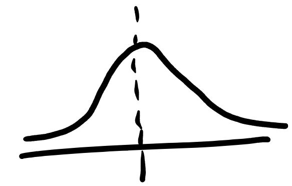
For $y > 0$, the cdf of V is given by

$$G(y) = P(V \le y) = P(Z^2 \le y) =$$

$$= P(-\sqrt{y} \le \underset{\substack{\uparrow \\ Z \sim N(0,1)}}{Z} \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-z^2/2}}_{} dz$$

$$\left\{ \begin{array}{c} Z^2 \le y, \; y \ge 0 \\ \Longleftrightarrow \\ -\sqrt{y} \le Z \le \sqrt{y} \end{array} \right.$$

pdf of $Z \sim N(0,1)$ is an even function



even

because pdf is even

$$= 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz$$

To compute the pdf of V, differentiate the cdf of V, use the fundamental theorem of Calculus to get

$$g(y) = \frac{d}{dy} G(y) = \frac{d}{dy}\, 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

$$= 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \cdot \underbrace{\frac{d}{dy}\sqrt{y}}_{\text{from chain rule}} =$$

$$= 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-y/2} \cdot \frac{1}{2\sqrt{y}} = \underbrace{\boxed{\frac{1}{\sqrt{\pi}\, 2^{1/2}}\, y^{1/2-1} e^{-y/2}}}_{\frac{1}{\Gamma(\alpha)\cdot\theta^{\alpha}} \cdot y^{\alpha-1} e^{-y/\theta}} \quad r = 1$$

pdf of $\chi^2(1)$ $\longrightarrow$

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

with $\theta = 2$, $\alpha = \frac{1}{2}$

$$g(y) = \frac{d}{dy} G(y) = \frac{d}{dy} 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

$$= 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \cdot \underbrace{\frac{d}{dy}\sqrt{y}}_{\text{from Chain rule}} =$$

$$= \cancel{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y/2} \cdot \frac{1}{\cancel{2}\sqrt{y}} = \boxed{\frac{1}{\sqrt{\pi}\, 2^{1/2}}\, y^{1/2 - 1}\, e^{-y/2}}$$

$$\text{pdf of } \chi^2(1) \rightarrow \underbrace{\frac{1}{\Gamma(\alpha) \cdot \theta^{\alpha}} \cdot y^{\alpha - 1} e^{-y/\theta}}$$

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

with $\theta = 2$, $\alpha = \frac{1}{2}$

$r = 1$

FTC :

Review

1

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$$

$$\frac{d}{dx}\int_a^{g(x)} f(t)\,dt = f(g(x))\cdot g'(x)$$

Why? Say f has antiderivative F (so that $F' = f$)

$$\int_a^{g(x)} f(t)\,dt = F(g(x)) - F(a) \Rightarrow \frac{d}{dx}\int_a^{g(x)} f(t)\,dt = \frac{d}{dx}\,F(g(x)) - \cancel{F(a)}$$

$$= f(g(x))\cdot g'(x)$$

FTC Review 2

$$\frac{d}{dx}\int_{g_1(x)}^{g_2(x)} f(t)\,dt = f(g_2(x)) \cdot g_2'(x) - f(g_1(x)) \cdot g_1'(x)$$

Why

$$\frac{d}{dx}\left[\int_a^{g_2(x)} f(t)\,dt + \int_{g_1(x)}^{a} f(t)\,dt\right] =$$

$$= \frac{d}{dx}\left[\int_a^{g_2(x)} f(t)\,dt - \int_a^{g_1(x)} f(t)\,dt\right]$$