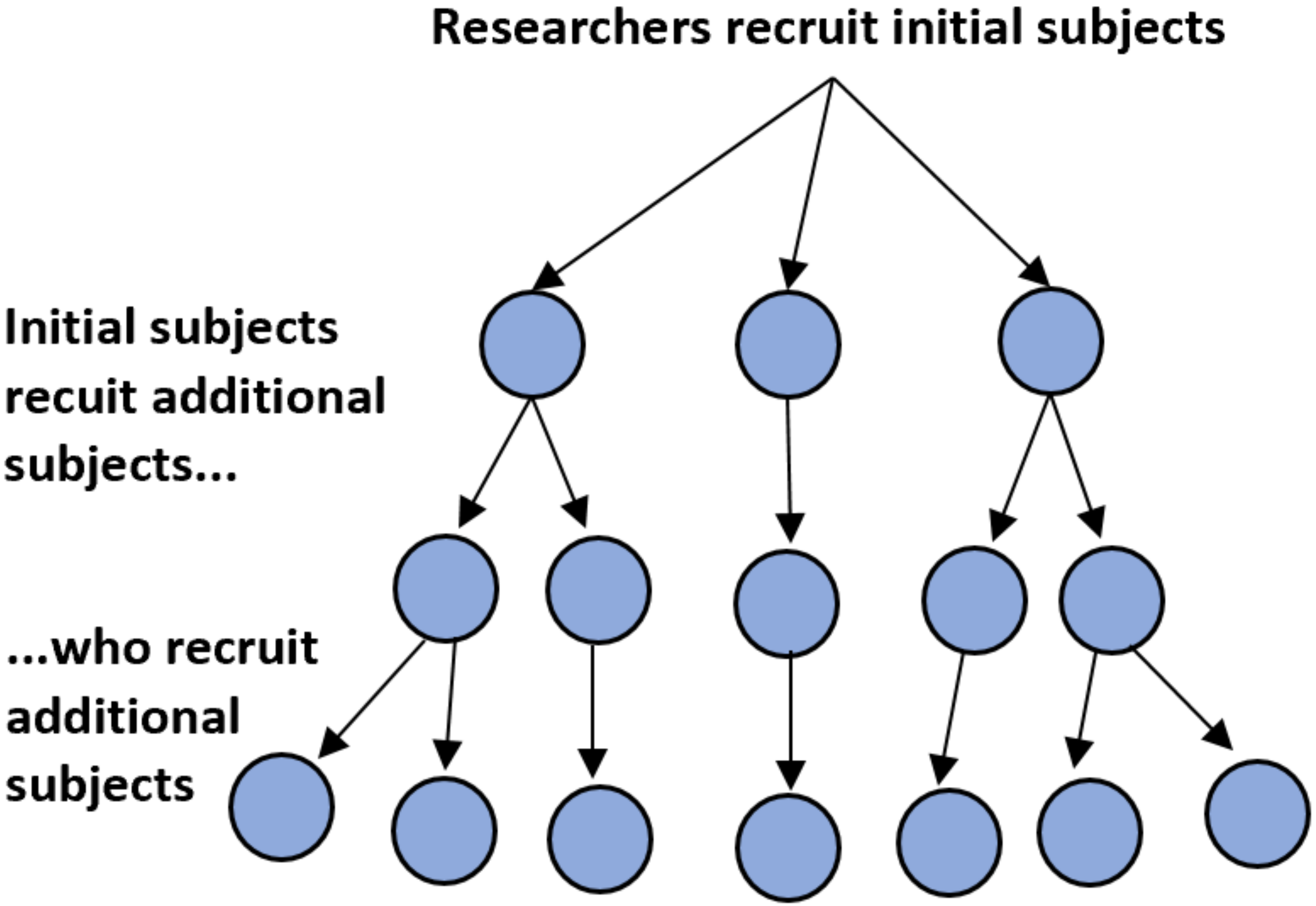
Sampling Distributions

Mainak Thakur

07.09.2020

Snowball Sampling



Ref: https://www.statology.org/snowball-sampling/

Sampling Distributions

- Statistic is a random variable. The value of a statistic may vary from sample to sample
- Imagine collecting a number of samples (each of size n) from a population and calculating the statistic for each sample. This series of realizations of statistic may be plotted in a histogram to see its frequency distribution.
- For large number of samples (each of size n), this distribution is called the sampling distribution of the statistic.
- Sampling distribution: mean, s.d, moments of higher orders
- Standard deviation of statistic is known as standard error of the statistic

Question

- What is the expectation and the standard error of the statistic sample mean if n i.i.d samples are taken from a population with mean μ and s.d. σ?

Question

- What is the expectation and the standard error of the statistic sample mean if n i.i.d samples are taken from a population with mean μ and s.d. σ?

Ans: Expectation = μ

Standard Error = $\dfrac{\sigma}{\sqrt{n}}$

Four fundamental distributions derived from normal

- Standard normal distribution (Z)
- Chi-square (χ^2) distributiom
- t distribution
- F distribution

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Standard Normal Distribution (Z)

Def^n: Standard normal variable (s.n.v.) is a normal variable with mean 0 and s.d. 1.

pdf of Z: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty$

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Theorem and Results

$\boxed{\text{Theorem:}}$ If X is normally distributed with mean μ and variance σ^2. Then $Y = a + bX$, where $b \neq 0$, is also normally distributed with mean $a + b\mu$ and variance $b^2\sigma^2$.

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

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

Converse is also true.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Notation

z_α : the value of Z s.t. $P(Z > z_\alpha) = \alpha$.

z_α is called the upper α-point (or the upper $100\alpha\%$-point) of Z.

Symmetry of Z distribution about 0

ensures: $z_{1-\alpha} = -z_\alpha$. $\left[\because P(Z > z_\alpha) = P(Z < -z_\alpha) = \alpha\right]$

$\left[P(Z > -z_\alpha) = 1-\alpha = P(Z > z_{1-\alpha})\right]$

$z_{1-\alpha}$: the lower α-point.

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

.45

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Graph

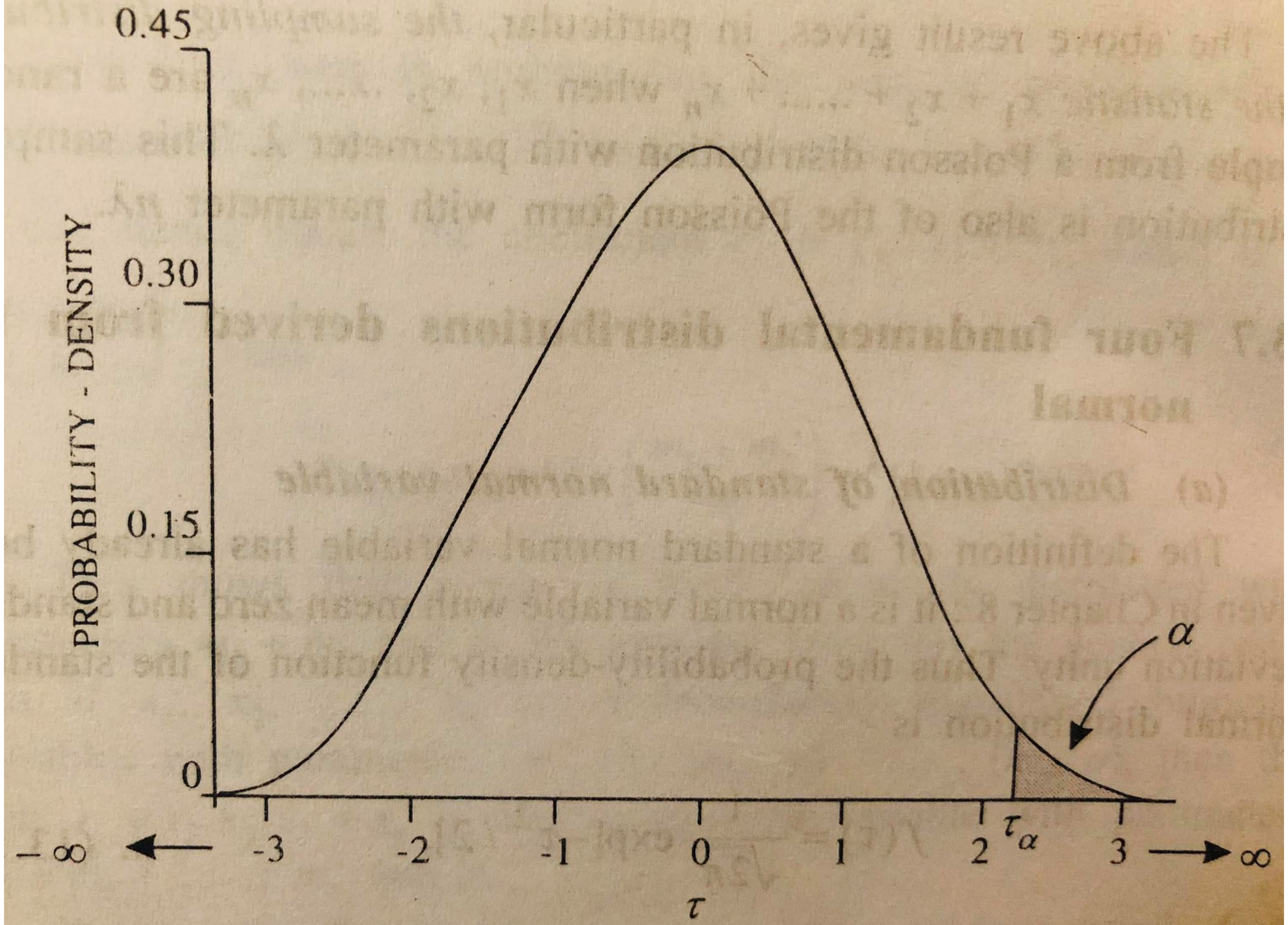


Fig. 13.1 Distribution of a standard normal variable.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Chi-square Distribution

NO

DATE

χ^2 distribution

Let $X_1, X_2, \ldots, X_\nu$ be ν mutually independent s.n.v. Then $\sum_{i=1}^{\nu} X_i^2$ is known to have χ^2 dist^n with ν degrees of freedom (or df$=\nu$).

pdf: $$f(\chi^2) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} \cdot \exp\left[-\chi^2/2\right] (\chi^2)^{\frac{\nu}{2}-1},$$

$$0 < \chi^2 < \infty.$$

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Properties

Mean: ν

Variance: 2ν.

$\chi^2(1) \rightarrow$ mean $= 1$

variance $= 2$

The term "degrees of freedom" refers to the no. of independent s.n.v. present.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

https://web.ma.utexas.edu/users/mks/M358KInstr/tDistandDF.pdf

Example

Ex: Take i.i.d. random samples $X_1, X_2, \ldots, X_k$ from some $N(\mu, \sigma^2)$ population.

$\frac{X_i - \mu}{\sigma}$ is s.n.v.

$\frac{X_1 - \mu}{\sigma}, \ldots, \frac{X_k - \mu}{\sigma}$ are independent

$$\Rightarrow \left(\frac{X_1 - \mu}{\sigma}\right)^2 + \ldots + \left(\frac{X_k - \mu}{\sigma}\right)^2 \sim \chi^2(k)$$

Question: What can you say about the distribution of

$$\left(\frac{X_1 - \bar{X}}{\sigma}\right)^2 + \ldots + \left(\frac{X_n - \bar{X}}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$\left[S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \right]$$

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta
https://web.ma.utexas.edu/users/mks/M358KInstr/tDistandDF.pdf

Results

Result: $\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$$

Results! $\bar{X}$ and S are independent r.v.s

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta
https://web.ma.utexas.edu/users/mks/M358KInstr/tDistandDF.pdf

Particular Case

$\boxed{n=2} \rightarrow \quad X_1 - \bar{X} = \dfrac{X_1 - X_2}{2}$

$$X_2 - \bar{X} = \frac{X_2 - X_1}{2}$$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2}\left[2\left(\frac{X_1 - X_2}{2}\right)^2\right]$$

$$= \left(\frac{X_1 - X_2}{\sqrt{2}\,\sigma}\right)^2$$

$$E(X_1 - X_2) = 0$$

$$\text{var}(X_1 - X_2) = \text{Var}(X_1) + \text{Var}(X_2)$$

$$= 2\sigma^2$$

$$n=2 \Rightarrow \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(1).$$

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta
https://web.ma.utexas.edu/users/mks/M358KInstr/tDistandDF.pdf

Properties

<u>Results</u>: Y_1 and Y_2 are two independent r.v.s distributed as χ^2 with d.f. ν_1 and ν_2 respectively.

$\Rightarrow Y_1 + Y_2 \sim \chi^2(\nu_1 + \nu_2)$.

Notation

$\chi^2_{\alpha,\nu}$: value of $\chi^2(\nu)$ for which

$$P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha.$$

$\chi^2_{\alpha,\nu}$: upper α point of the χ^2 dist^n with

$df = \nu$,

$\chi^2_{1-\alpha,\nu}$: lower α point

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Graph

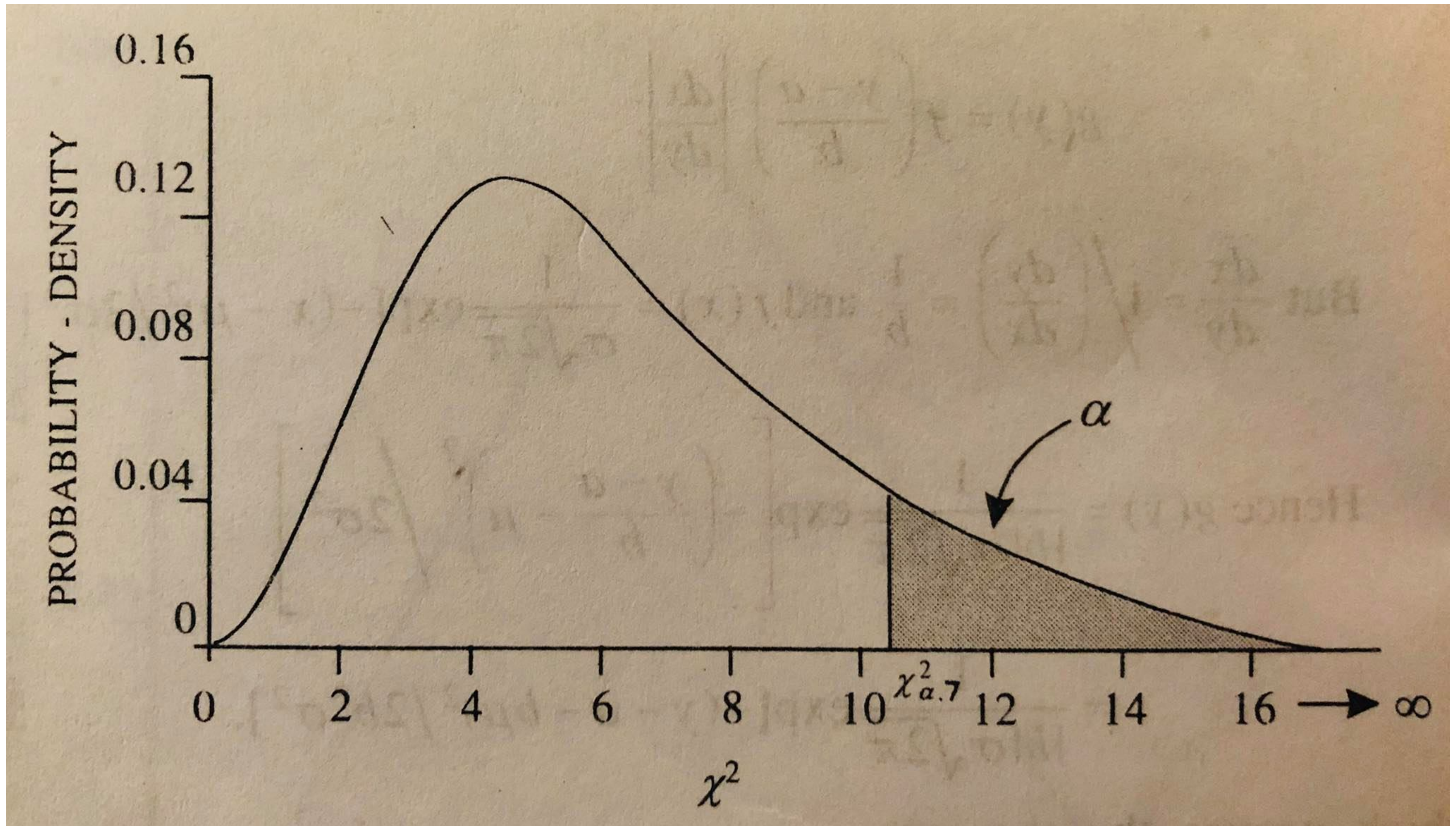
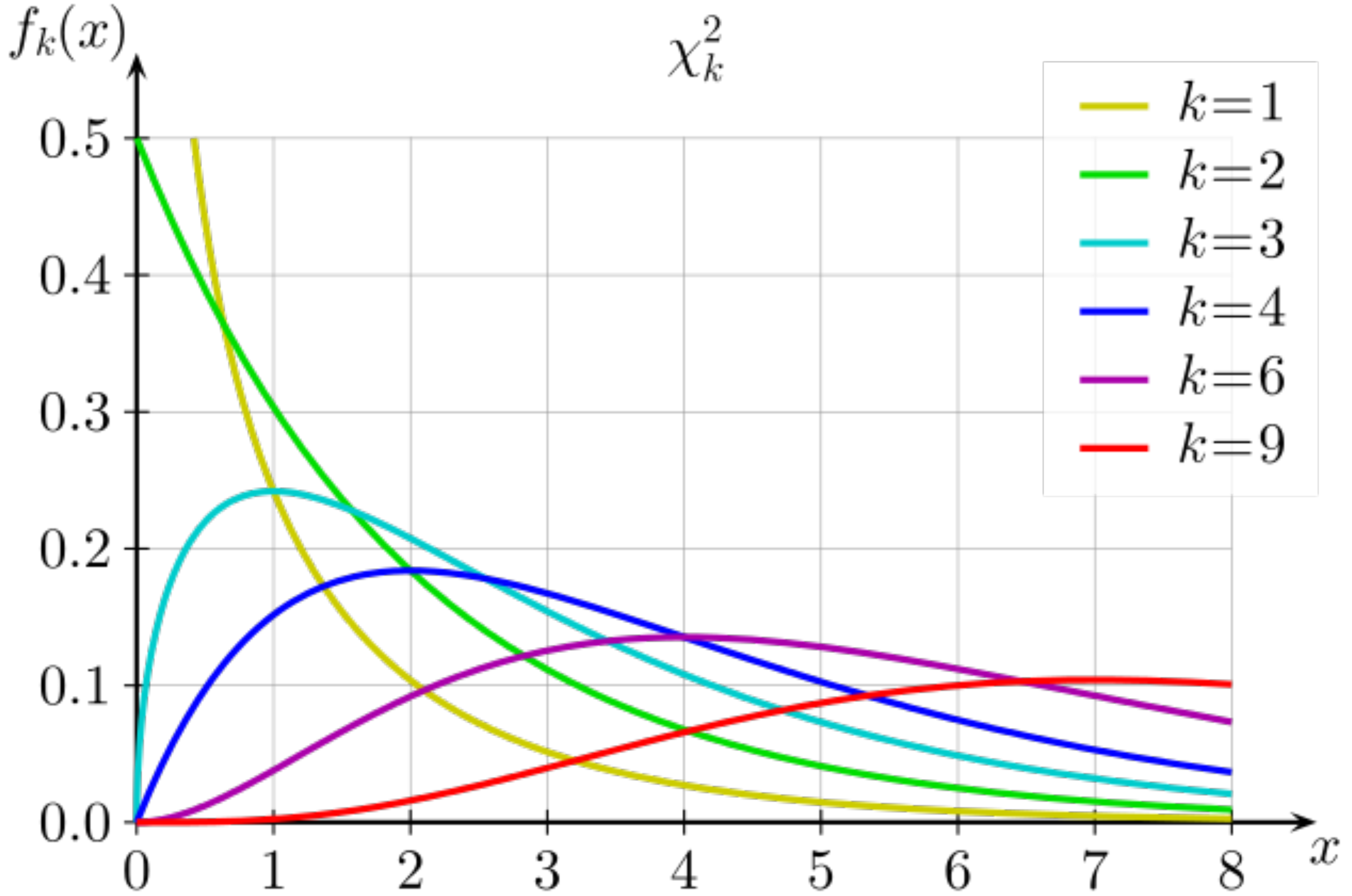


Fig. 13.2 χ^2 distribution with 7 degrees of freedom ($\nu = 7$).



Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta
wikipedia