LAB: RELATED DISTRIBUTIONS

Lab 5.1

CONTENTS

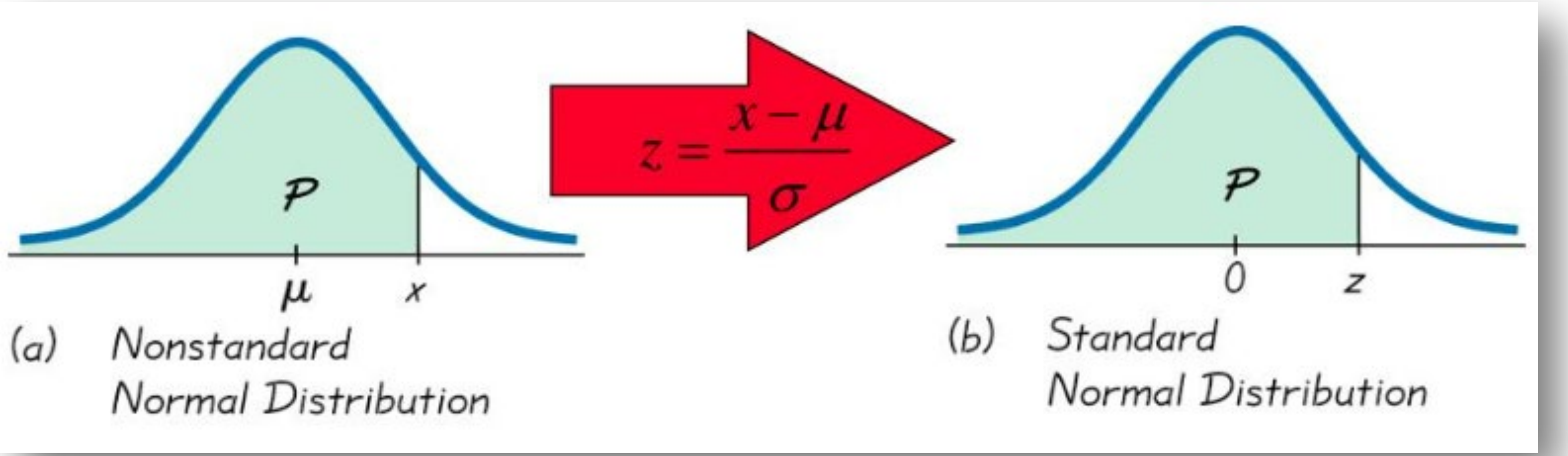
Standard Normal distribution

Chi-square distribution

F-distribution

t distribution

Standard Normal Distribution



PDF:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < +\infty$$

CDF:

$$\Phi(z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz, -\infty < z < +\infty$$

Chi-square Distribution

Chi-square goodness of fit test

Chi-square test of independence

Imagine taking a random sample of a standard normal distribution (Z). If you squared all the values in the sample, you would have the chi-square distribution with k = 1.

$$x_1^2 = (z)^2$$

Now imagine taking samples from two standard normal distributions (Z1 and Z2). If each time you sampled a pair of values, you squared them and added them together, you would have the chi-square distribution with k = 2.

$$x_2^2 = (z_1)^2 + (z_2)^2$$

More generally, if you sample from k independent standard normal distributions and then square and sum the values, you'll produce a chi-square distribution with k degrees of freedom.

$$x_k^2 = (z_1)^2 + (z_2)^2 + \dots + (z_k)^2$$

$$Y \sim x_n^2$$

PDF of chi-square distribution

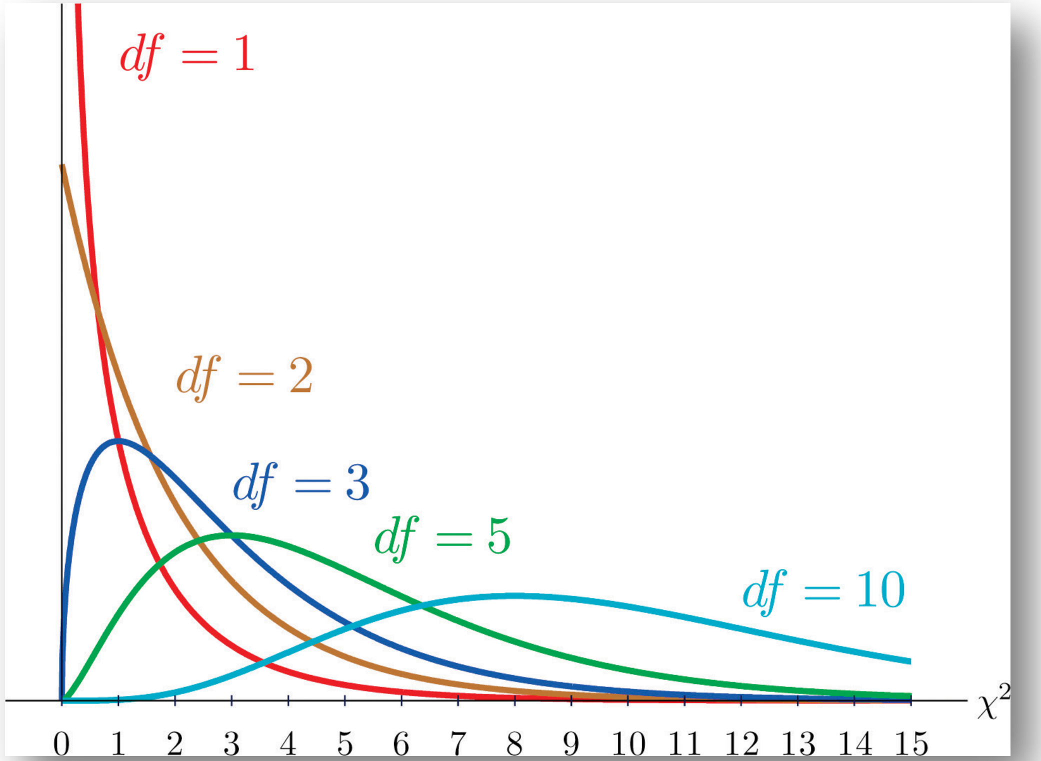
$$f(x) = \begin{cases} \dfrac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} & \text{for } x > 0, \\ 0 & \text{for } x \le 0 \end{cases}$$

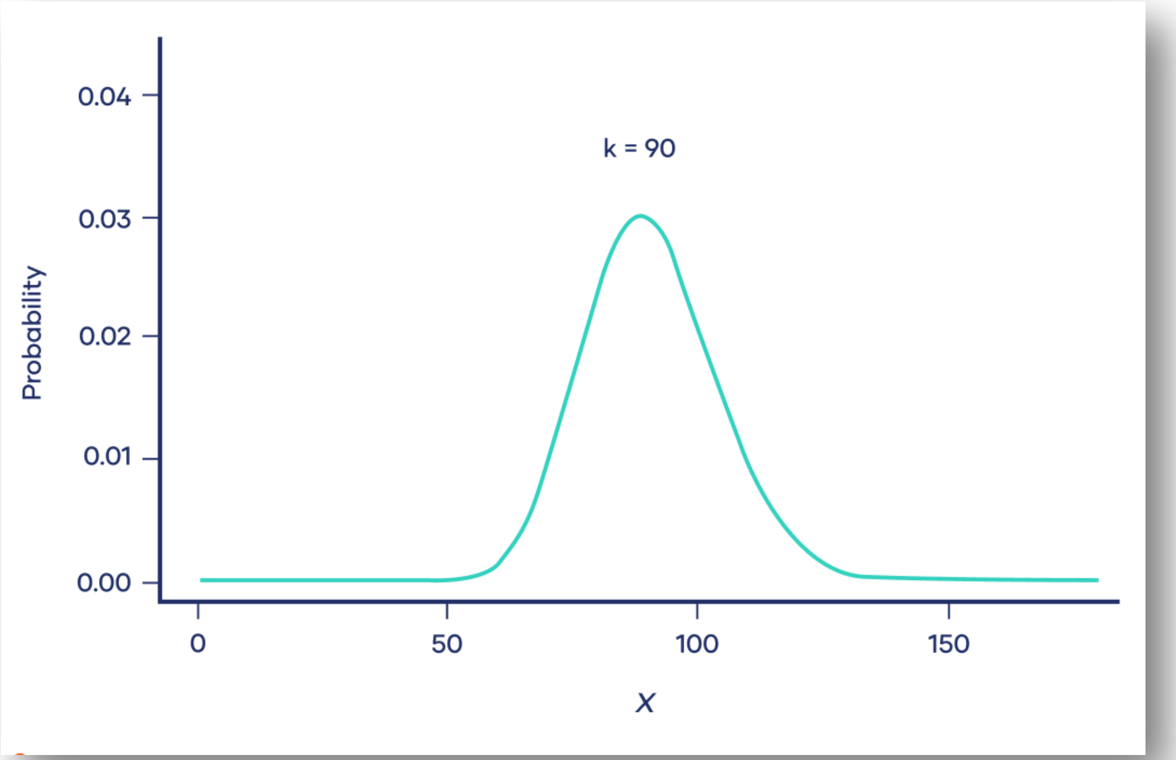
where:

k = Positive Integer Denoting Degrees of Freedom,

Karl Pearson（1857 ~ 1936）

Γ = Gamma Function





exponential distribution is a special chi-square distribution

Properties of chi-square distribution

(1) If $Y \sim x_n^2$, then E(Y)=n, and D(Y)=2n

Prove

$$Y = x_1^2 + x_2^2 + x_3^2 + \ldots.. + x_n^2$$

$$Ex_i^2 = Dx_{\mathrm{i}} + (Ex_{\mathrm{i}})^2 = 1$$

$$EY = E(\sum_{i=1}^{n} x_{\mathrm{i}}^2) = \sum_{i=1}^{n} Ex_{\mathrm{i}}^2 = n$$

$$D(X) = E\{[X - E(X)]^2\}$$
$$= E\{X^2 + E^2(X) - 2X E(X)\}$$
$$= E(X^2) + E^2(X) - 2E(X)\cdot E(X)$$
$$= E(X^2) - E^2(X)$$
$$\therefore E(X^2) = D(X) + E^2(X)$$

$$
\begin{aligned}
Var(x) &= E((x - E(x))^2) \\
&= E(x^2 - 2xE(x) + E(x)^2) \\
&= E(x^2) - E(2xE(x)) + E(E(x)^2) \\
&= E(x^2) - \sum 2E(x)xf(x) + \sum E(x)^2 f(x) \\
&= E(x^2) - 2E(x)^2 + E(x)^2 \sum f(x) \\
&= E(x^2) - 2E(x)^2 + E(x)^2 \\
&= E(x^2) - E(x)^2
\end{aligned}
$$

$$because: Dx_i^2 = Ex_i^4 - (Ex_i^2)^2 = 3 - 1 = 2$$

$$thus: DY = D(\sum_{i=1}^{n} x_i^2) = \sum_{i=1}^{n} Dx_i^2 = 2n$$

$$EX_i^4 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = 3$$

Properties of chi-square distribution

(2) If $Y_1 \sim x_m^2, Y_2 \sim x_n^2$, and Y1 and Y2 are independent

Then $Y_1 + Y_2 \sim x_{m+n}^2$

Properties of chi-square distribution

(3) If $X \sim N(\mu,\sigma^2),(X_1,\ldots,X_n)$ are randomly selected from population X

The distribution of $Y = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$

$$because\ X \sim N(\mu,\sigma^2),\ so\ \frac{x_i-\mu}{\sigma} \sim N(0,1)$$

According to the definition of chi-square distribution

$$Y=\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 \sim x_n^2$$

Properties of chi-square distribution

(4)

$$x_n^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \approx \sum_{i=1}^{n} \frac{(x_i - \bar{X})^2}{\sigma^2}$$

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{X})^2}{n-1}$$

$$\sum_{i=1}^{n} (x_i - \bar{X})^2 = (n-1) * S^2$$

$$x^2 \approx \sum_{i=1}^{n} \frac{(n-1) * S^2}{\sigma^2}$$

$$\sim x_{n-1}^2$$

F Distribution

ANOVA

if $X \sim x_n^2, Y \sim x_m^2$. **X and Y are independent.**

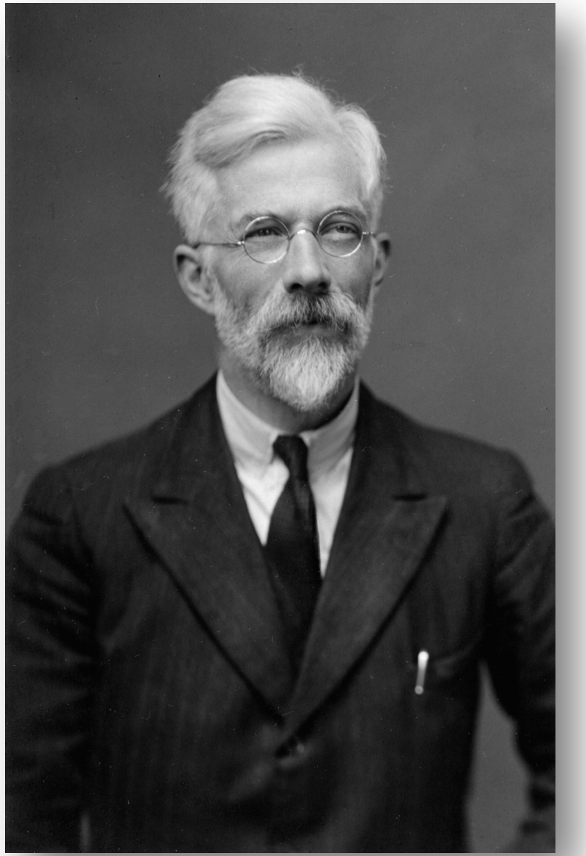
$$F = \frac{X/n}{Y/m}, then$$

$$F \sim F_{n,m}, or F \sim F(n,m)$$

The PDF of an *F* random variable with r1 numerator degrees of freedom and **r2** denominator degrees of freedom is:

$$f(w) = \frac{(r_1/r_2)^{r_1/2}\Gamma[(r_1+r_2)/2]w^{(r_1/2)-1}}{\Gamma[r_1/2]\Gamma[r_2/2][1+(r_1 w/r_2)]^{(r_1+r_2)/2}}$$

over the support w≥0.



Ronald A. Fisher

Properties of an F-distribution

(1) If F~F(n, m), then 1/F~F(m, n)

Prove

$$X \sim x_n^2, Y \sim x_m^2$$

$$F_{n,m} = \frac{X/n}{Y/m}, then$$

$$1/F = \frac{Y/m}{X/n} \sim F_{m,n}$$

Properties of an F-distribution

(2) If $T \sim t_n$, then $T^2 \sim F(1, n)$

Prove According to the definition of t distribution

$$T = \frac{x}{\sqrt{Y/n}}$$

Of which, $X \sim N(0,1)$, $Y \sim X^2_n$, and X and Y are independent

$$T^2 = \frac{x^2}{Y/n} = \frac{x^2/1}{Y/n}$$

Of which, $X^2 \sim X^2_1$

t Distribution

t-test

If $X \sim N(0,1)$, $Y \sim X^2{}_n$, and X and Y are independent

$$T = \frac{x}{\sqrt{Y/n}}$$

Then T follows a t distribution with df=n, we write

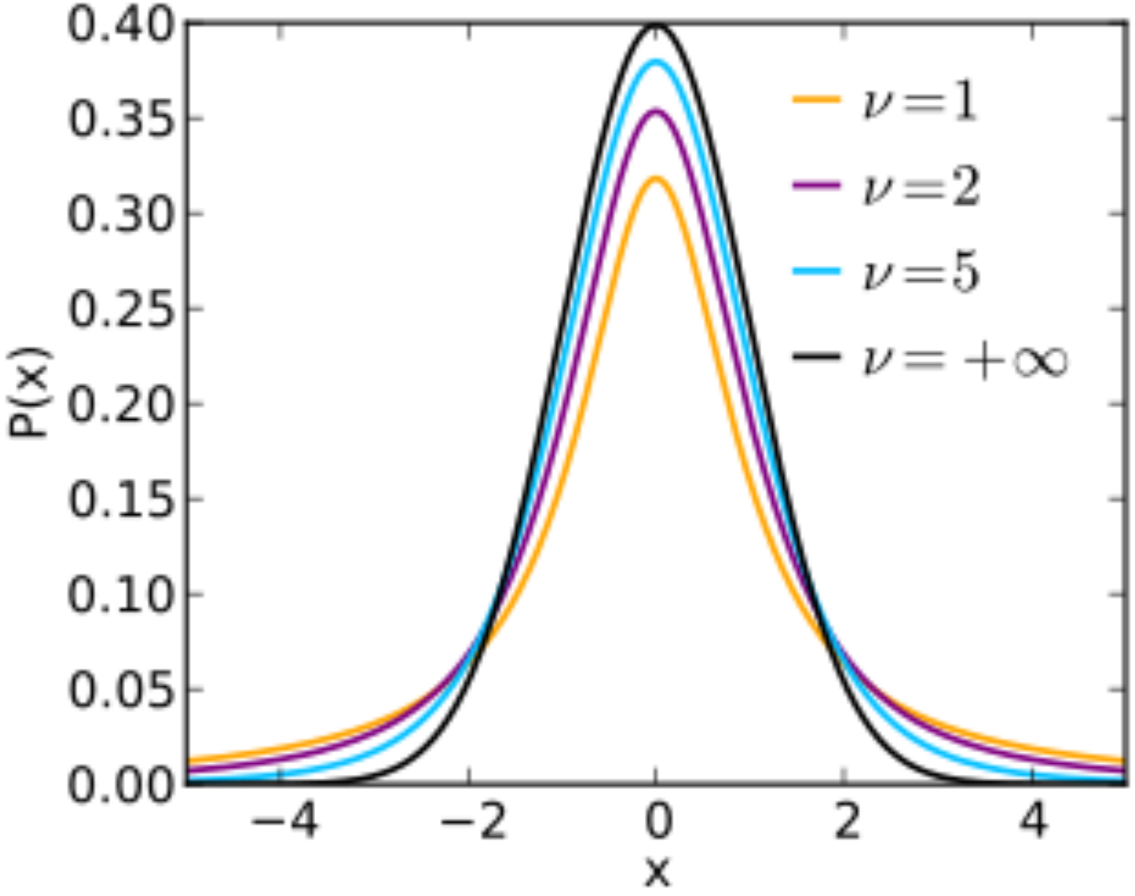
$$T \sim t_n$$

$$f(t) = \frac{\Gamma\left[(\nu+1)/2\right]}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

$$-\infty < \nu < +\infty$$

William Sealy Gosset

Student in 1908



Properties of t-distribution

(1) If n=1, then it follows a Cauchy distribution

(2) If n>2, then ET=0, DT=$\frac{n}{n-2}$

(3) If n becomes large, the t distribution converges to an N(0, 1) distribution.

(4) t distributions have heavier tails than N(0,1).

Exercise, Using R

- 1. plot standard normal curve, and return the cumulative probability of the standard normal distribution at 1.96;
- 2. construct a $\chi 2$ distribution (df=1), make a histogram of x2;
- 3. construct a t-distribution (with df=5) from the normal and $\chi 2$ distributions, generate a histogram for the data;
- 4. randomly select 6 values from N(0,1) then replete 10000 times. Plot a histogram of its sample mean.