AMS 131
20 Jul 18

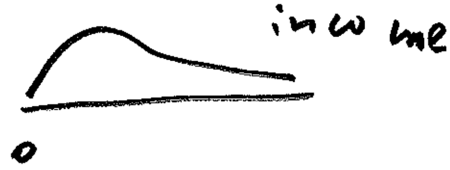
1

this time: catalogue of discrete
next time: & continuous distributions

read: DS ch. 5

"Normal" dist.

name: Karl Pearson (1880)




Gaussian

↑

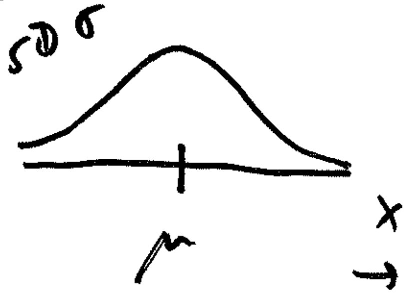
K. F. Gauss (~1780)

first: A. de Moivre (1710)

5050

| | | |
|---|---|---|
| 1 | 100 | 101 |
| 2 | 99 | |
| ⋮ | | |
| 50 | 51 | |

} 50




$X \sim \text{Normal}(\mu, \sigma^2)$

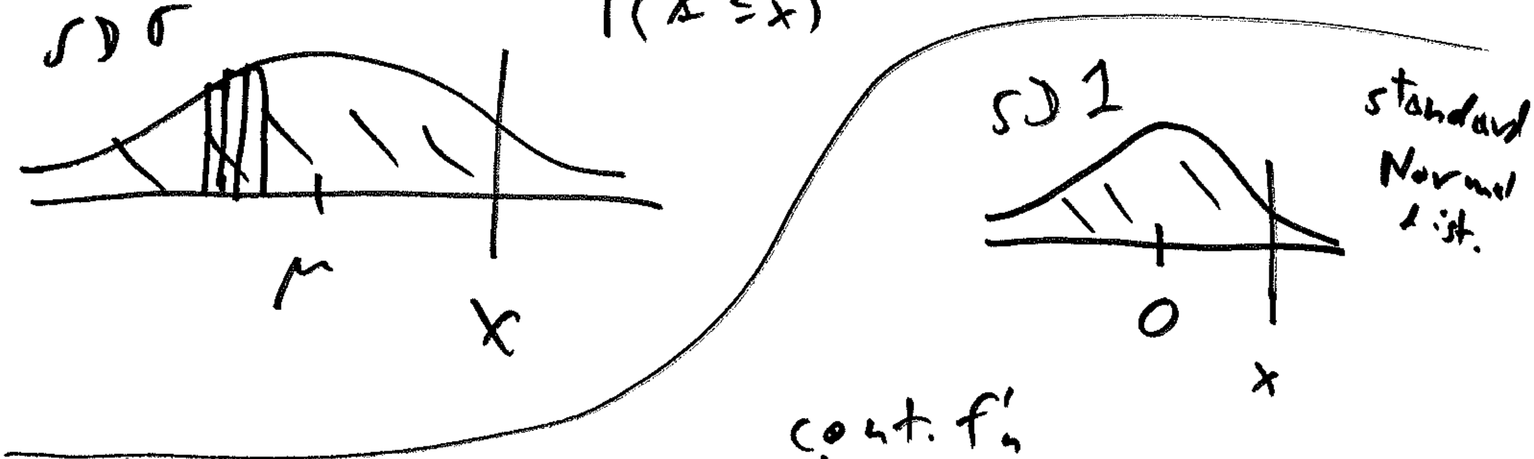
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

(9.48)

$\boxed{\text{cont.}}$ RV $\quad X \sim N(\mu, \sigma^2)$

$$F_X(x) = \boxed{\text{cont.}}\text{ f'n} = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt$$

$$= P(X \le x)$$




CDF of standard Normal Dist.:

cont. f'n

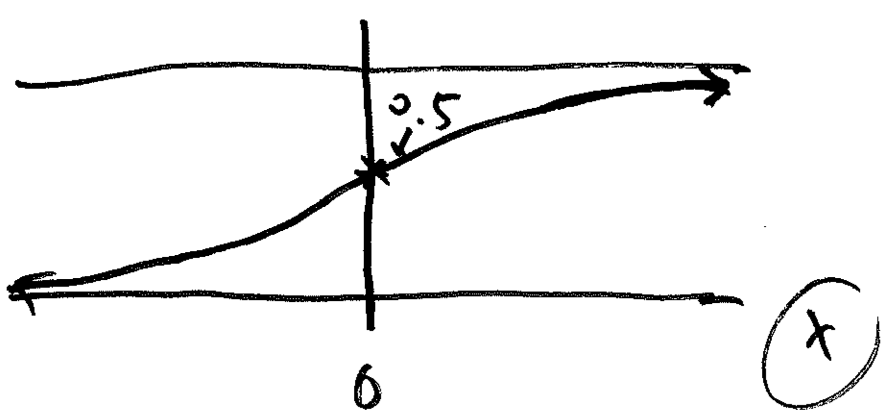
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right] dt$$

capital phi

numerical int.

or ratio of polynomials approx

$$\frac{ax^3 + bx^2 + cx + d}{ex^4 + fx^3 + gx^2 + hx + i}$$



**Empirical Rule** Pretty much no matter what the PDF (histogram) of a continuous univariate RV looks like,

if you start at mean & go [1, 2, 3] either way,

you will usually capture [about $\frac{2}{3}$ (68%), most (95%), almost all (99.7%)] of the dist.

(R)

density → **d**norm ← PDF $f_X(x)$

**p**norm ← CDF $F_X(x)$

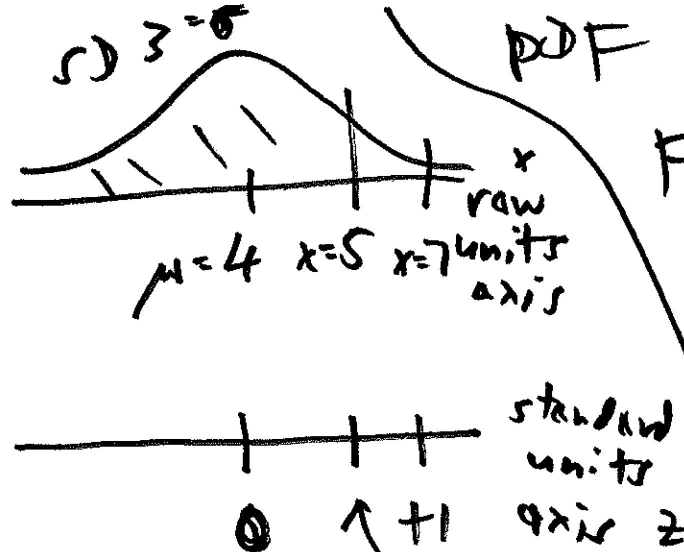
quantile **q**norm ← inverse CDF

random draws **r**norm ←

$F_X(x) = p$

$x = F_X^{-1}(p)$

(10.35)




$\frac{5-4}{3}$

PDF of $X$

$$F_X(5) = P(X \le 5)$$

$$= P\left( \frac{X-\mu}{\sigma} \le \frac{5-\mu}{\sigma} \right)$$

$$= \Phi\left( \frac{5-\mu}{\sigma} \right) = \Phi\left( \frac{5-4}{3} \right)$$

$$= \Phi\left( \frac{1}{3} \right) \doteq 0.63$$

$$\boxed{z = \frac{x-\mu}{\sigma}}$$

standardization = converting to standard units

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \qquad \begin{bmatrix} y_1 \checkmark \\ y_2 \checkmark \\ X \end{bmatrix}$$

✓ = free to vary

mean 4

$$4 = \frac{y_1 + y_2 + y_3}{3}$$

a data set with n obs. has only (n − 1) degrees of freedom for measuring spread

$$\left.\begin{array}{c}(X_i \mid \mu, \sigma^2) \overset{IID}{\sim} N(\mu, \sigma^2) \\ (i = 1, \dots, n)\end{array}\right\} \rightarrow \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \overset{ID}{=} \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{\cancel{n}\mu}{\cancel{n}} = \mu$$

$$V(\bar{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$$

$$\overset{I \; ID}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \overset{ID}{=} \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

so $\boxed{SD(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}}$ ← important formula give $\underline{X}_1, \dots, \underline{X}_n \overset{IID}{\sim}$ (population mean $\mu$, SD $\sigma$)

I estimate that $\mu$ is around $\bar{X}_n$, give or take $\frac{\sigma}{\sqrt{n}}$