AMS131
26 Aug 19

1

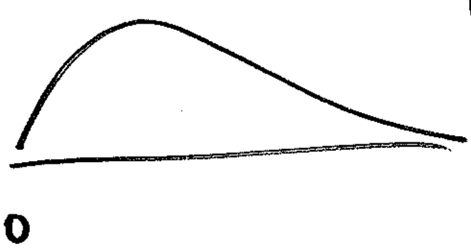
this time: Lognormal, Gamma, Beta, Multinomial, Bivariate Normal

next time: large random samples

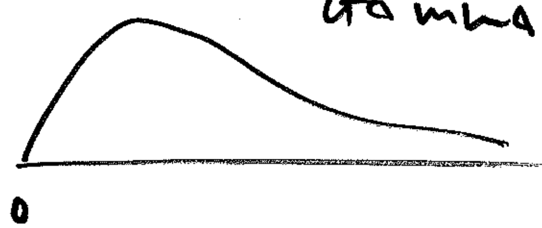
Quiz 8 due tomorrow night

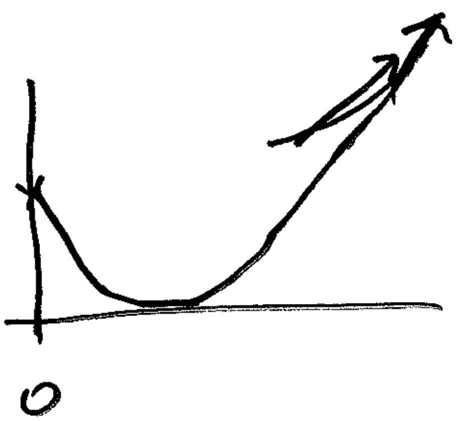
THT 3 due Sun night 1 Sep 2019

Lognormal



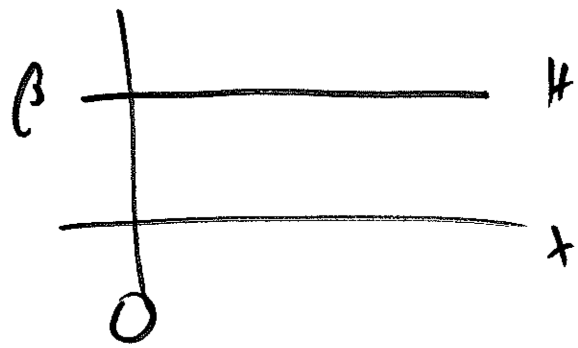
Gamma

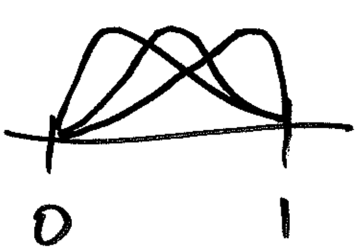




$H(x)$

age →



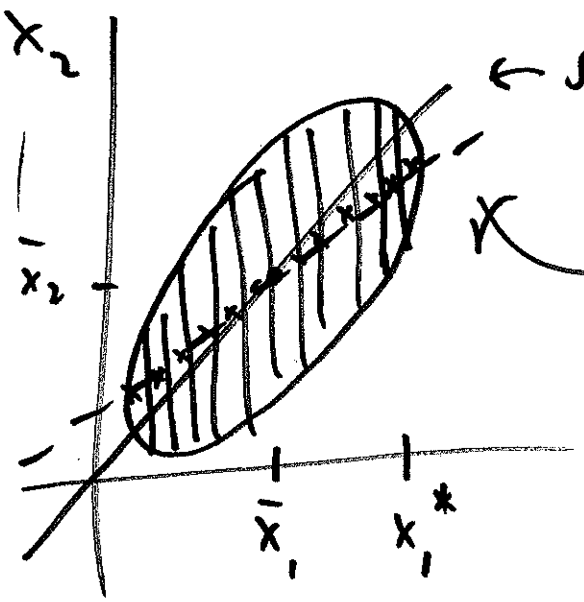


unknown

$\theta = P(\sim)$

(9.50)

(10.47)



← SD line : slope $\frac{s_{x_2}}{s_{x_1}}$

regression line

slope $r \cdot \frac{s_{x_2}}{s_{x_1}} = \hat{\beta}_1$

$$\hat{x}_2 = \hat{\beta}_0 + \hat{\beta}_1 x_1$$

$$\bar{x}_2 = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1$$

$$\hat{\beta}_0 = \bar{x}_2 - \hat{\beta}_1 \bar{x}_1$$

$$\hat{x}_2 = \bar{x}_2 - \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_1 x_1$$

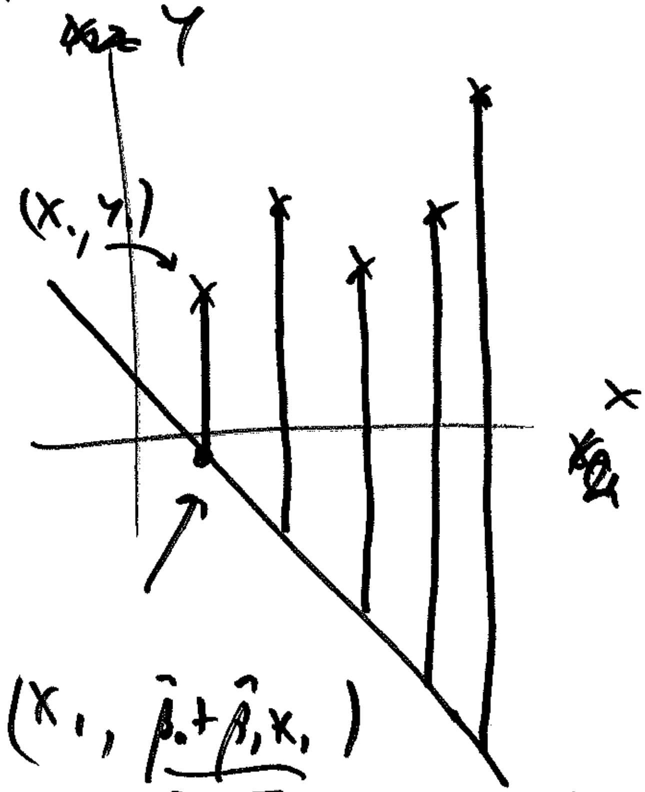
empirical

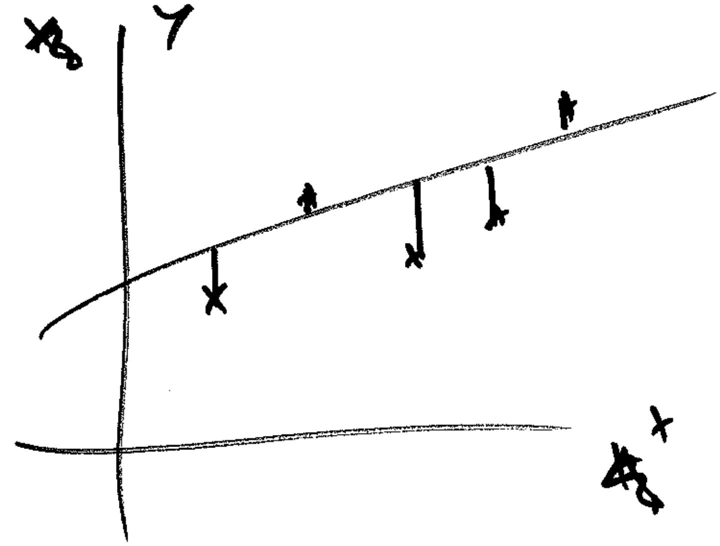
$$= \bar{x}_2 + \hat{\beta}_1 (x_1 - \bar{x}_1) = \bar{x}_2 + r \frac{s_{x_2}}{s_{x_1}} (x_1 - \bar{x}_1)$$

$$\hat{x}_2 \nearrow$$

theoretical story

$$E(X_2 \mid X_1 = x_1) = \hat{x}_2 = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$$





Gauss: find best line for predicting x_2 from x_1

$|\cdot|$ Laplace
$(\ )^2$ Gauss

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right]^2 = g(\hat{\beta}_0, \hat{\beta}_1)$$

find $(\hat{\beta}_0, \hat{\beta}_1)$ to minimize $g(\hat{\beta}_0, \hat{\beta}_1)$

→ least squares line:

= regression line

$$\hat{\beta}_1 = r\,\frac{s_y}{s_x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

pop. $\quad$ $X_i \sim$ cont. $\quad$ sample

intuition: unbiased

$\bar{X}_n$ is a good estimate of μ

$N = ?$

IID $\rightarrow$ $\begin{array}{c} X_1 \\ \vdots \\ X_n \end{array} \left(\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right) \updownarrow n$

mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\left(\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \right)$

mean $\mu = E(X_i) \leftarrow$ (1)

SD $\sigma = \sqrt{V(X_i)}$

pop. PDF

TS

pop. $\xrightarrow{\text{case 1}}$ sample

$\xleftarrow{\text{case 2}}$

probability

case 1 pop. known, random sample not yet taken (unknown)

case 2 random sample known, pop. unknown: statistical inference (harder)

induction (inference)

$\longleftarrow$

pop $\qquad$ sample

general $\longrightarrow$ particular

whole $\longrightarrow$ part $\quad$ (easier)

deduction (probability)

pop

sample

repeated-sampling dataset

all possible $\bar{X}_n$

$\cancel{\text{IID}}$

$n = 25$

mean $\bar{X}_n = ?$ (ex. 10.4)

10.4

9.9

$\vdots$

$M \to \infty$

mean μ = known (ex. 10)

SD $\sigma = 2 < \infty$

IID

n

mean $\bar{X}_n = ?$ (ex. 9.9)

instead

long run mean — by WLLN $E(\bar{X}_n) = \mu = 10$

long run SD — by WLLN $SE(\bar{X}_n) = \sqrt{V(\bar{X}_n)} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{25}} = 0.4$

long run PDF

long run PDF