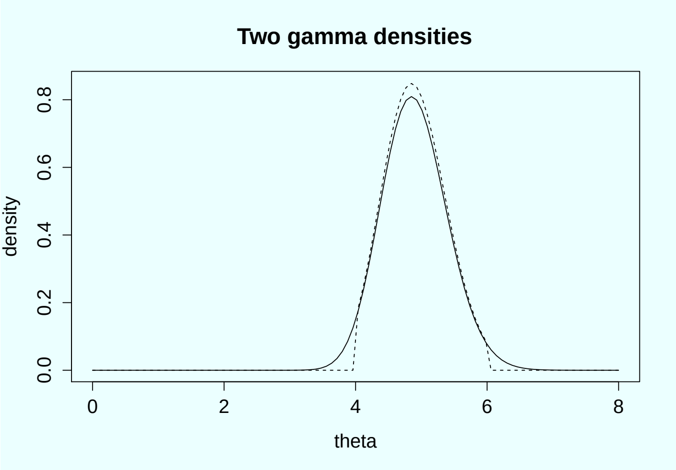
Lab 3

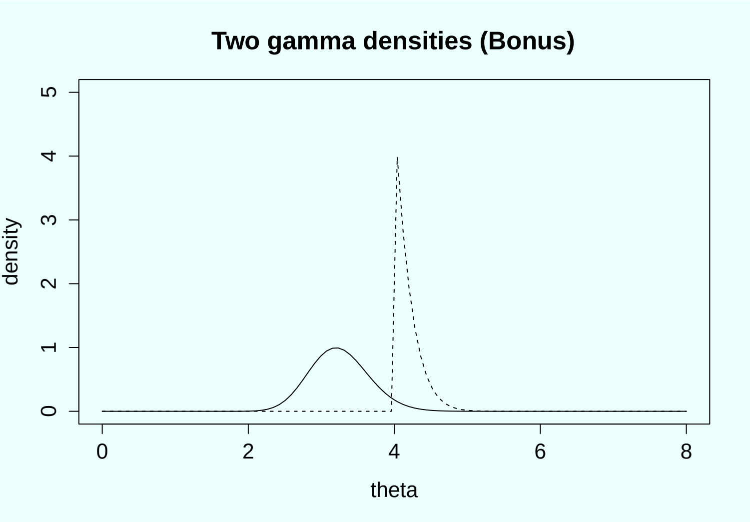
January 29, 2013

HW 2: Posterior densities



(Untruncated) [3.978, 5.916],   (Truncated) [4.123, 5.791]

Posterior densities with the second data set



Prior predictive distribution

- $\pi(\theta)$: a prior for a parameter θ.
- $L(x \mid \theta)$: a data generating function.
- Prior predictive distribution:

$$\pi(x) = \int L(x \mid \theta)\pi(\theta)d\theta, \tag{1}$$

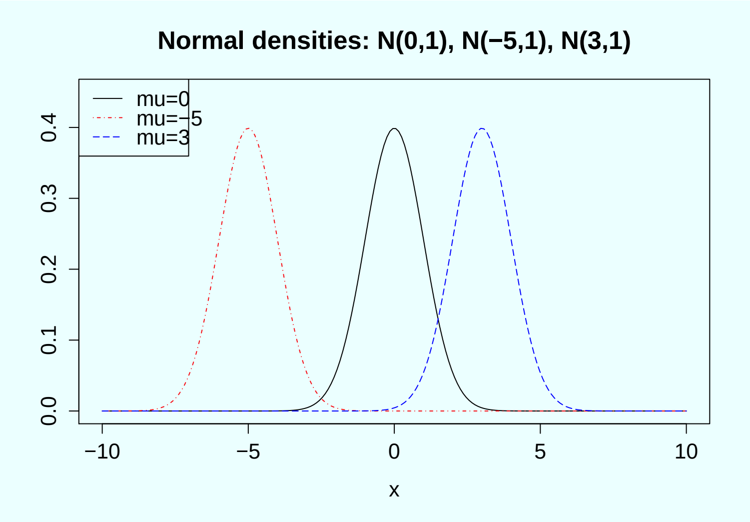
- Generate random samples from $x_i \sim \pi(x)$ for $i = 1, \ldots, S$.
  1. Generate $\theta^{(i)} \sim \pi(\theta)$.
  2. Generate $x^{(i)} \sim L(x \mid \theta^{(i)})$.

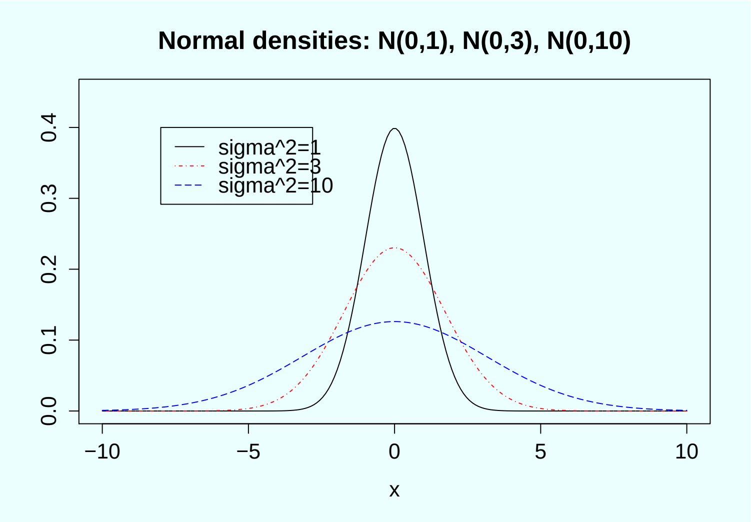
Normal distribution

- A random variable Y is said to be normally distributed $Y \sim N(\theta, \sigma^2)$ if the pdf is given by

$$f(y \,|\, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\theta)^2\right\}, \quad -\infty < y < \infty. \tag{2}$$

- θ: mean, σ^2: variance.
- Symmetric about θ, the mode, median and mean are equal to θ.
- If $Y \sim N(\theta, \sigma^2)$, then $aY \sim N(a\theta, a^2\sigma^2)$.
- If $Y \sim N(\theta, \sigma^2)$ and $X \sim N(\mu, \tau^2)$, then $aY + bX \sim N(a\theta + b\mu, a^2\sigma^2 + b^2\tau^2)$.


Normal densities: N(0,1), N(−5,1), N(3,1)
mu=0
mu=−5
mu=3
0.0
0.1
0.2
0.3
0.4
−10
−5
0
5
10
x



Normal densities: N(0,1), N(0,3), N(0,10)
sigma^2=1
sigma^2=3
sigma^2=10
0.0
0.1
0.2
0.3
0.4
−10
−5
0
5
10
x

Central limit theorem

Let $\{X_n, n \geq 1\}$ be i.i.d. random variables with $E(X_n) = \theta$ and $V(X_n) = \sigma^2 < \infty$. If $\bar{X}_n = \sum_{i=1}^{n} X_i/n$, then

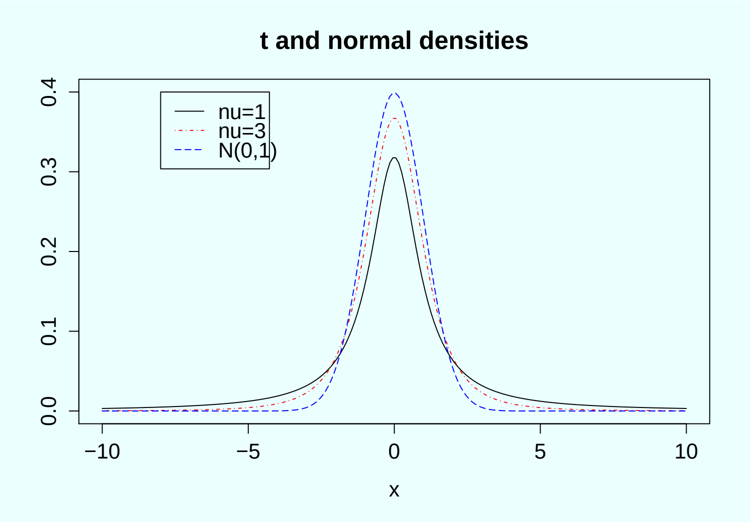
$$\sqrt{n}\frac{\bar{X}_n - \theta}{\sigma} \to N(0, 1). \tag{3}$$

Student's t distribution

- A random variable X follows t distribution if the pdf is

$$f(x \mid \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{4}$$

  where ν is the number of degrees of freedom.
- $\nu \to \infty$ corresponds to normal distribution.


t and normal densities
nu=1
nu=3
N(0,1)
0.0
0.1
0.2
0.3
0.4
−10
−5
0
5
10
x

Kolmogorov-Smirnov test

- KS test checks the sample is generated from the reference distribution.
- KS statistic relies on the empirical distribution for i.i.d. $\{X_i, i = 1, \ldots, n\}$,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x}, \tag{5}$$

  where $I_{X_i \leq x}$ is an indicator function.
- The statistic is

$$D_n = \sup_x |F_n(x) - F(x)|, \tag{6}$$

  and $D_n \to 0$ if the null hypothesis is true.

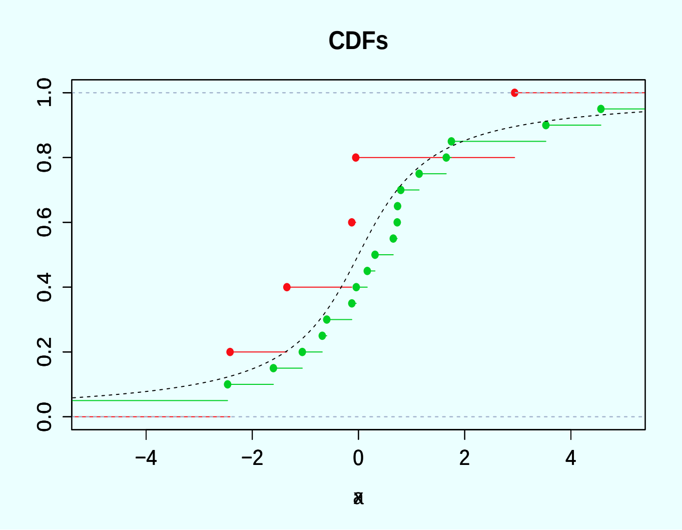


Figure: CDFs of t dist with $\nu = 1$ (red: #=5, green: #=20).

Kolmogorov-Smirnov test

- $x = \text{rt}(100, 1)$
  ks.test(x,pt,1)
  D = 0.115, p-value = 0.1667
  $\Rightarrow$ Fail to reject the hypothesis.
- $y = \text{rnorm}(100, 0, 3)$
  ks.test(y, pt, 1)$ p
  0.0359
  $\Rightarrow$ Reject the hypothesis.