Final Exam Review

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

Transformation of Variables

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

Change of Variable

Given a random variable x, with probability distribution function $f_X(x|\theta)$, we often would like to know the probability distribution of a random variable y, that is a function $y(x)$ of x, $y=y(x)$.

Change of Variable

Let y=y(x) be a one-to-one transformation

with inverse transformation x=x(y) .

Then, if $f_X(x\,|\,\theta)$ is the PDF of x, the PDF of y can be found as

$$f_Y(y\,|\,\theta) = f_X(x(y)\,|\,\theta) \times |\, J(x \to y)\,|$$

where $J(x \to y) = \dfrac{dx(y)}{dy}$.

Suppress PDF subscripts.

Change of Variable

Not one-to-one

Let $y=y(x)$ be a not one-to-one transformation,
(i.e. $y=x^2$, then $x_1(y)=+\sqrt{y}$ and $x_2(y)=-\sqrt{y}$.)

We can still perform the change of variable by breaking up the transformation into pieces that are1-to-1.

$$f_Y(y|\theta)=\sum_j f_X(x_j(y)|\theta)\times\left|\frac{dx_j(y)}{dy}\right|$$

i.e. $f_Y(y|\theta)=f_X(\sqrt{y}|\theta)\left|\frac{1}{2\sqrt{y}}\right|+f_X(-\sqrt{y}|\theta)\left|\frac{-1}{2\sqrt{y}}\right|$

Bivariate Transformation of Variables

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

Outline

- **Bivariate Continuous Distributions**
  Joint PDF, Conditional PDF, Marginal PDF

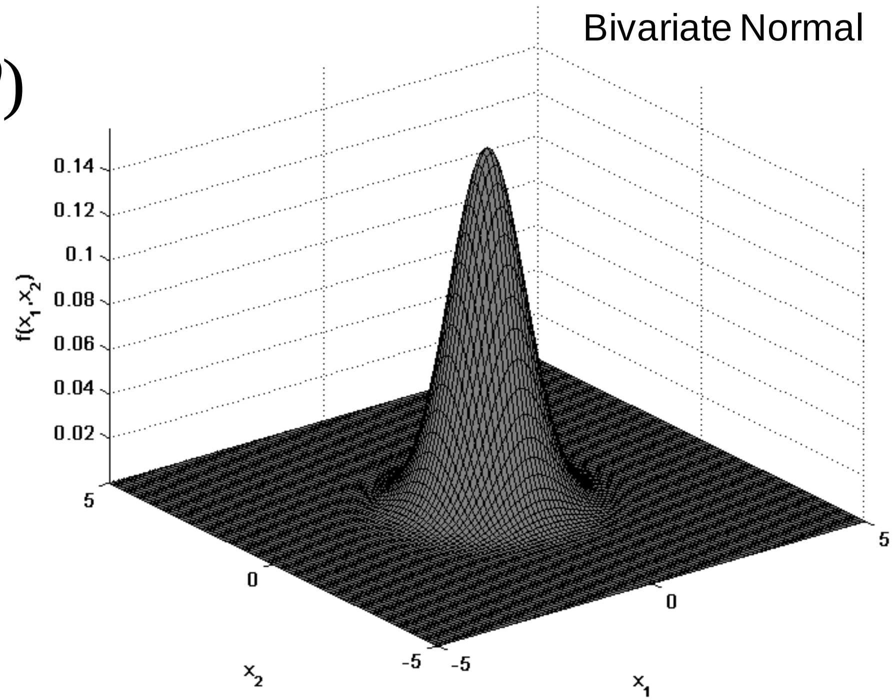
- **Bivariate Transformation of Variables**
  Sum of RVs

Bivariate Continuous Distributions

A bivariate (2D) PDF $f(x_1, x_2 \mid \theta)$ of two continuous random variables (x_1, x_2) depending upon parameters θ satisfies

1) $0 \le f(x_1, x_2 \mid \theta) \le 1, \quad \forall (x_1, x_2)$

2) $\iint_{x_1 x_2} f(x_1, x_2 \mid \theta) dx_1 dx_2 = 1$.



Bivariate Change of Variable

Given two continuous random variables, (x_1, x_2)

with joint probability distribution function $f_{X_1,X_2}(x_1, x_2 \mid \theta)$.

Let $\begin{pmatrix} y_1(x_1, x_2) \\ y_2(x_1, x_2) \end{pmatrix}$ be a transformation from (x_1, x_2) to (y_1, y_2)

with inverse transformation $\begin{pmatrix} x_1(y_1, y_2) \\ x_2(y_1, y_2) \end{pmatrix}$.

Bivariate Change of Variable

Then, the joint probability distribution function $f_{Y_1,Y_2}(y_1,y_2|\theta)$

of (y_1,y_2) can be found via

$$f_{Y_1,Y_2}(y_1,y_2|\theta)=f_{X_1,X_2}\big(x_1(y_1,y_2),x_2(y_1,y_2)|\theta\big)\times|J(x_1,x_2\to y_1,y_2)|$$

where $J(x_1,x_2\to y_1,y_2)=\begin{vmatrix}\dfrac{dx_1(y_1,y_2)}{dy_1} & \dfrac{dx_1(y_1,y_2)}{dy_2}\\ \dfrac{dx_2(y_1,y_2)}{dy_1} & \dfrac{dx_2(y_1,y_2)}{dy_2}\end{vmatrix}$.

Bivariate Change of Variable - Sum

Let x_1 have PDF $f_{X_1}(x_1 | \theta_1)$ and x_2 have PDF $f_{X_2}(x_2 | \theta_2)$,

then, the PDF of $y_1 = x_1 + x_2$ can be found via the

bivariate change of variable technique

$$f_{Y_1,Y_2}(y_1, y_2 | \theta) = f_{X_1,X_2}(x_1(y_1, y_2), x_2(y_1, y_2) | \theta) \cdot |J(x_1, x_2 \rightarrow y_1, y_2)|$$

with marginalization $f_{Y_1}(y_1 | \theta) = \int_{y_2} f_{Y_1,Y_2}(y_1, y_2 | \theta)\, dy_2$.

Bivariate Change of Variable - Sum

The joint PDF of (x_1, x_2) is

$$f_{X_1,X_2}(x_1, x_2 \mid \theta) = f_{X_1}(x_1 \mid \theta_1) f_{X_2}(x_2 \mid \theta_2).$$ (need not be independent)

Let $y_1 = x_1 + x_2$ and $y_2 = x_2$, then $x_1 = y_1 - y_2$ and $x_2 = y_2$.

$$J(x_1, x_2 \to y_1, y_2) = \begin{vmatrix} \frac{dx_1(y_1, y_2)}{dy_1} & \frac{dx_1(y_1, y_2)}{dy_2} \\ \frac{dx_2(y_1, y_2)}{dy_1} & \frac{dx_2(y_1, y_2)}{dy_2} \end{vmatrix} = 1$$

Bivariate Transformation of Variables
(continued)

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

Outline

- Distribution of Mean of RVs
  Normal and Uniform

- Central Limit Theorem (CLT)
  Uniform vs. Normal

Bivariate Change of Variable - Average

Talked about x_1 with PDF $f_{X_1}(x_1 \mid \theta_1)$, and x_2 with PDF $f_{X_2}(x_2 \mid \theta_2)$, then the PDF of $y_1 = x_1 + x_2$ can be found via the bivariate change of variable technique

$$f_{Y_1,Y_2}(y_1, y_2 \mid \theta) = f_{X_1,X_2}(x_1(y_1,y_2), x_2(y_1,y_2) \mid \theta) \times |J(x_1,x_2 \to y_1,y_2)|$$

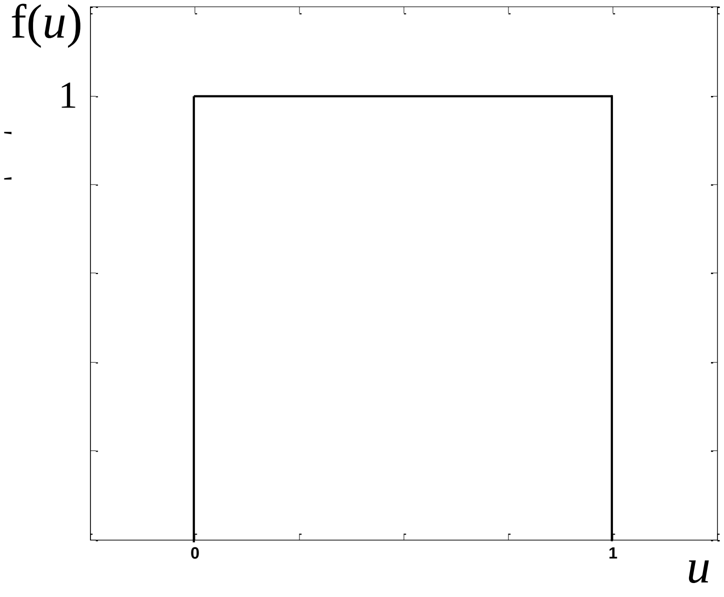
with marginalization $f_{Y_1}(y_1 \mid \theta) = \int_{y_2} f_{Y_1,Y_2}(y_1, y_2 \mid \theta)\, dy_2$.

Bivariate Change of Variable - Average

The change of variable technique can be used to find the distribution of the sum or average of two uniform RVs.

Let u_1~uniform(0,1) and u_2~uniform(0,1).
The joint PDF of (u_1,u_2) is

$$f(u_1,u_2)=\begin{cases}1 & \text{if } u_1\in[0,1] \text{ and } u_2\in[0,1]\\ 0 & \text{if } u_1\notin[0,1] \text{ or } u_2\notin[0,1]\end{cases}$$



Bivariate Change of Variable - Average

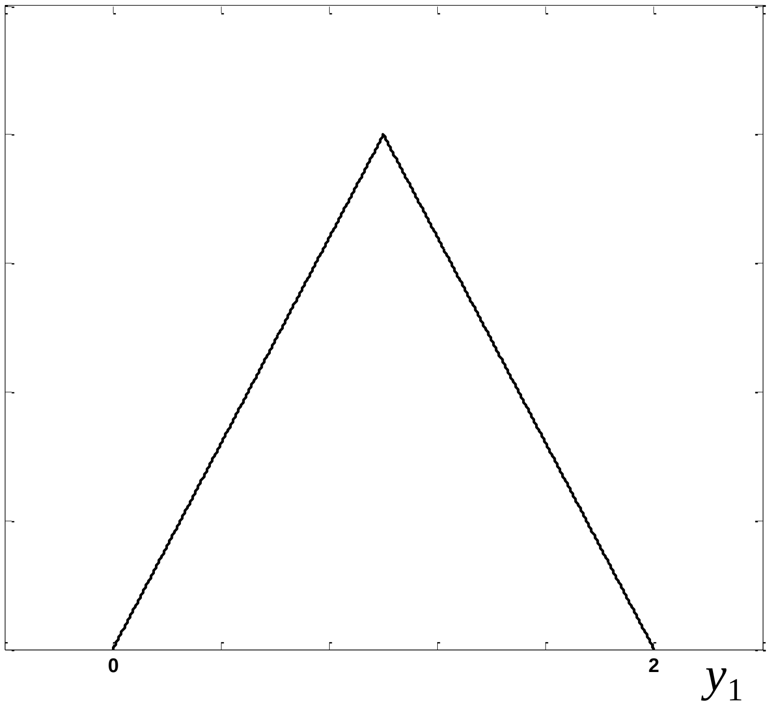
Using the change of variable technique

$$f_{Y_1,Y_2}(y_1,y_2\,|\,\theta)=f_{U_1,U_2}(u_1(y_1,y_2),u_2(y_1,y_2)\,|\,\theta)\times|J(u_1,u_2\rightarrow y_1,y_2)|$$

$$f_{Y_1}(y_1\,|\,\theta)=\int_{y_2} f_{Y_1,Y_2}(y_1,y_2\,|\,\theta)\,dy_2\ ,$$

the distribution of $y_1=u_1+u_2$ is

$$f(y_1)=\begin{cases} 0 & \text{if} \quad y_1<0 \\ y_1 & \text{if} \quad 0\le y_1<1/2 \\ 1-y_1 & \text{if} \quad 1/2\le y_1\le 1 \\ 0 & \text{if} \quad y_1>1 \end{cases}\ .$$



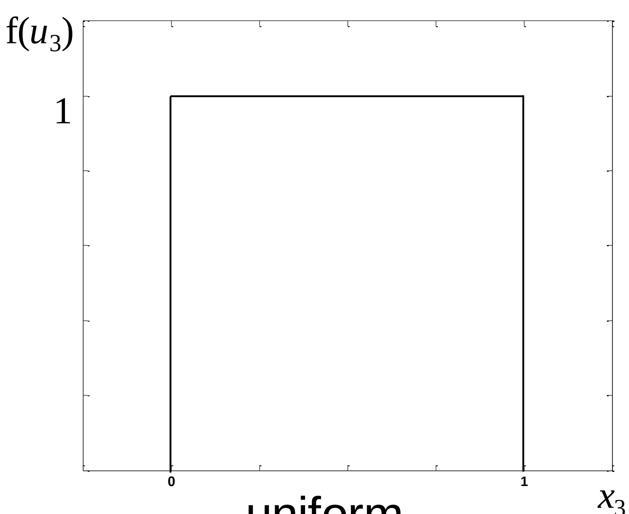
Bivariate Change of Variable - Average

This change of variable technique can be applied again, u_3~uniform(0,1), and the distribution of $y_3=y_1+u_3$ found

$$f_{Y_2,Y_3}(y_3,y_4\,|\,\theta)=f_{Y_3,Y_4}(y_1(y_3,y_4),u_3(y_3,y_4)\,|\,\theta)\times|J(y_1,u_3\to y_3,y_4)|$$

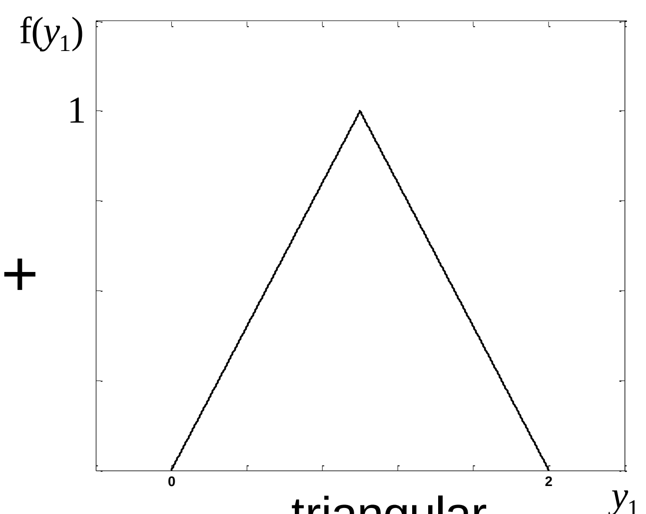
$$f_{Y_3}(y_3\,|\,\theta)=\int_{y_4} f_{Y_3,Y_4}(y_3,y_4\,|\,\theta)\,dy_4$$

y_4 is another variable not of interest for the bivariate change of variable



uniform

RV +



triangular

RV = ? RV

Homework.

Bivariate Change of Variable - Average

This change of variable technique can be repeated many times to determine the distribution of $y=(u_1+u_2+u_3+\ldots+u_n)/n$ where each u_i is uniform(a,b).

For large n, the mean becomes normally distributed with

$$\mu_{\bar{X}} = \mu, \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \;.$$

$$\mu = \frac{b-a}{2}$$

$$\bar{X} \sim N(\mu = \mu_{\bar{X}}, \sigma^2_{\bar{X}} = \sigma^2 / n)$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

The Central Limit Theorem (CLT)

Two continuous data examples, $\mu = 100$ and $\sigma = 57.73$.

Generate data $x_1,...,x_n$, for n=1, 2, 3, 4, 5, 15, 30, and 50.
Calculate $\bar{X}$, and repeat one million times.

Uniform distribution, [a=0,b=200].

$$f(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

Normal distribution, [μ=100,σ=57.7].

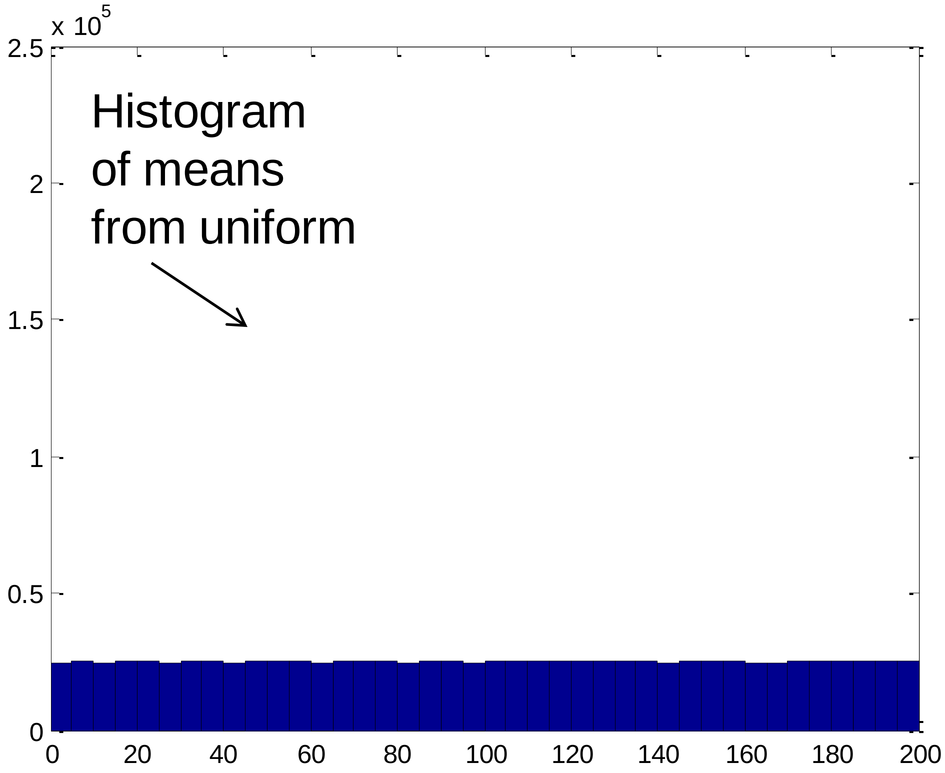
$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \qquad -\infty \leq x \leq \infty$$

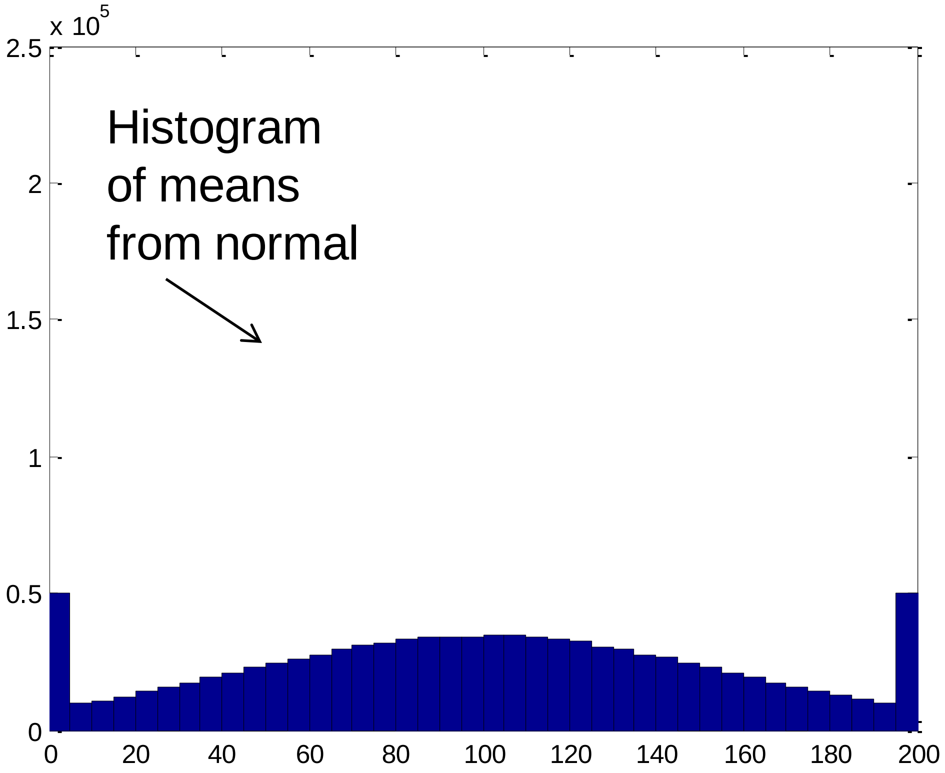
The Central Limit Theorem (CLT)

$\mu = 100$

$\sigma = 57.73$

n=1 1×10^6 means

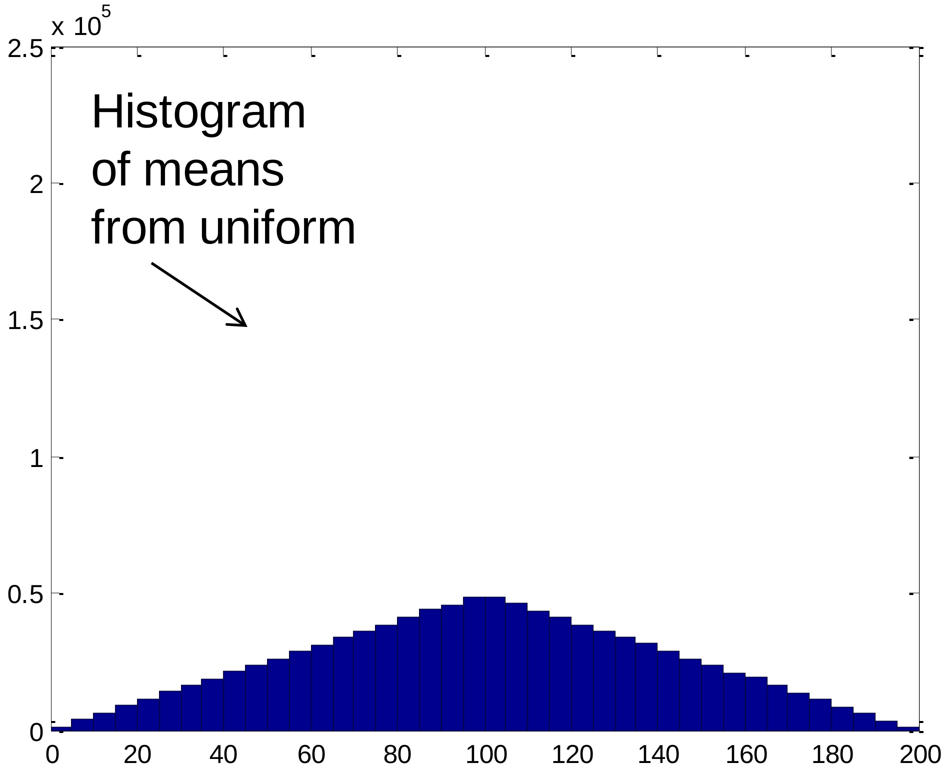


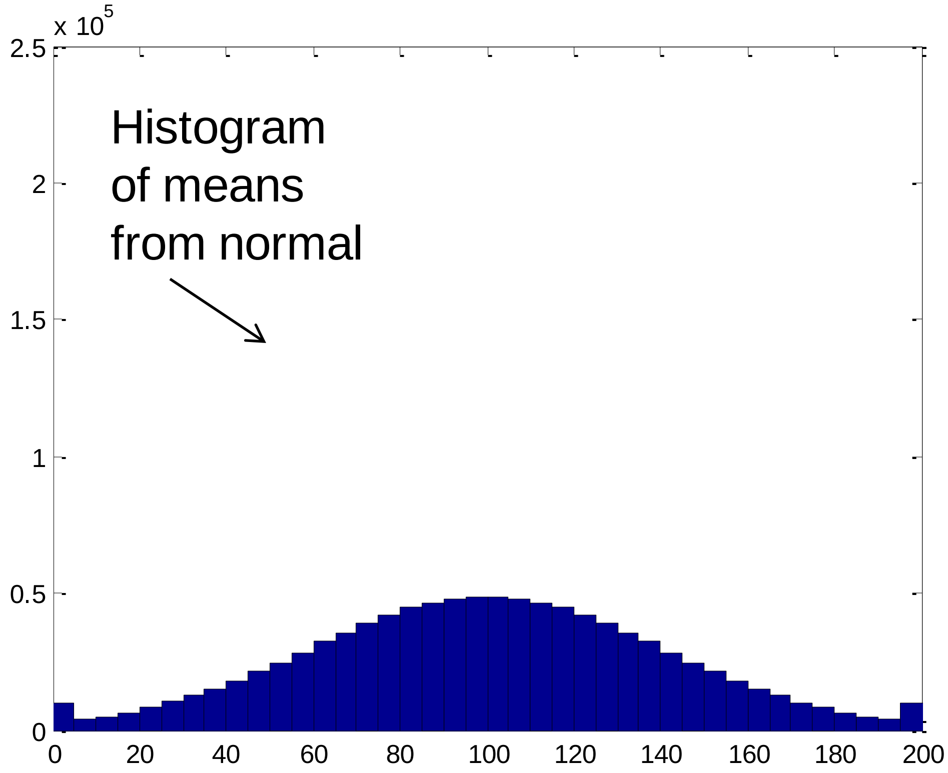


The Central Limit Theorem (CLT)

$$\mu = 100$$
$$\sigma = 57.73$$

n=2 1×10^6 means

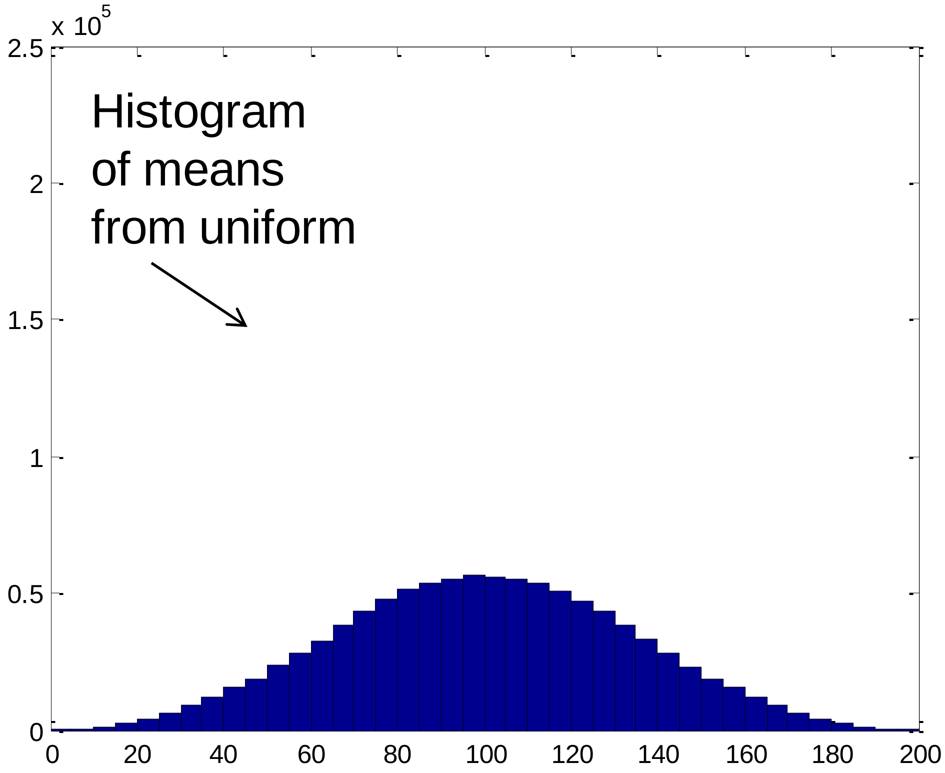


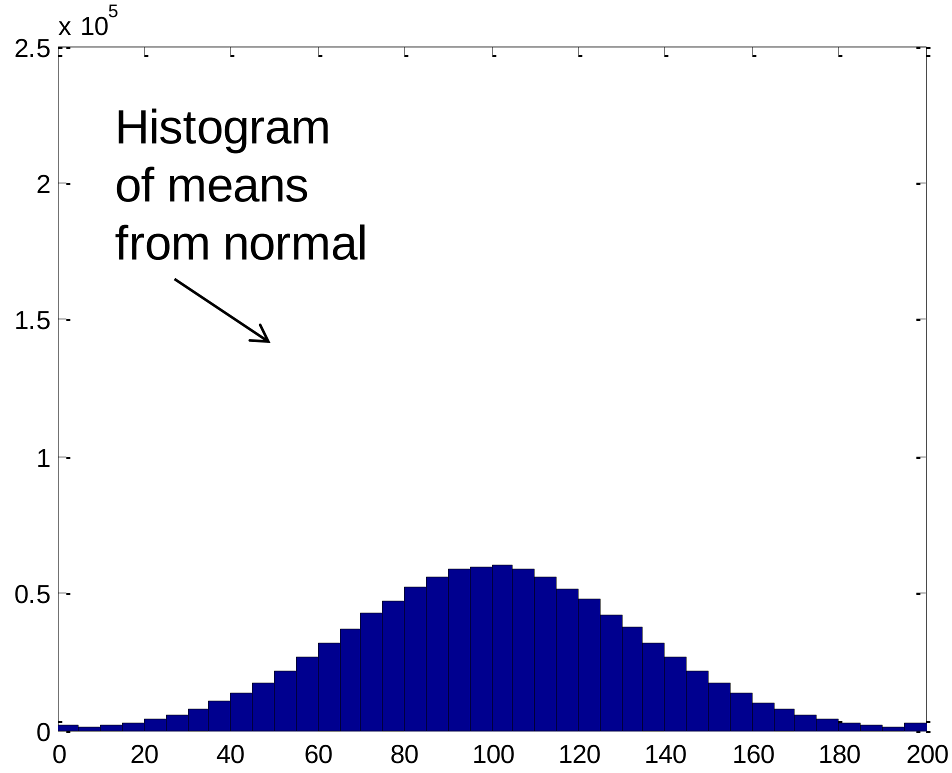


The Central Limit Theorem (CLT)

$$\mu = 100$$
$$\sigma = 57.73$$

n=3 1×10^6 means

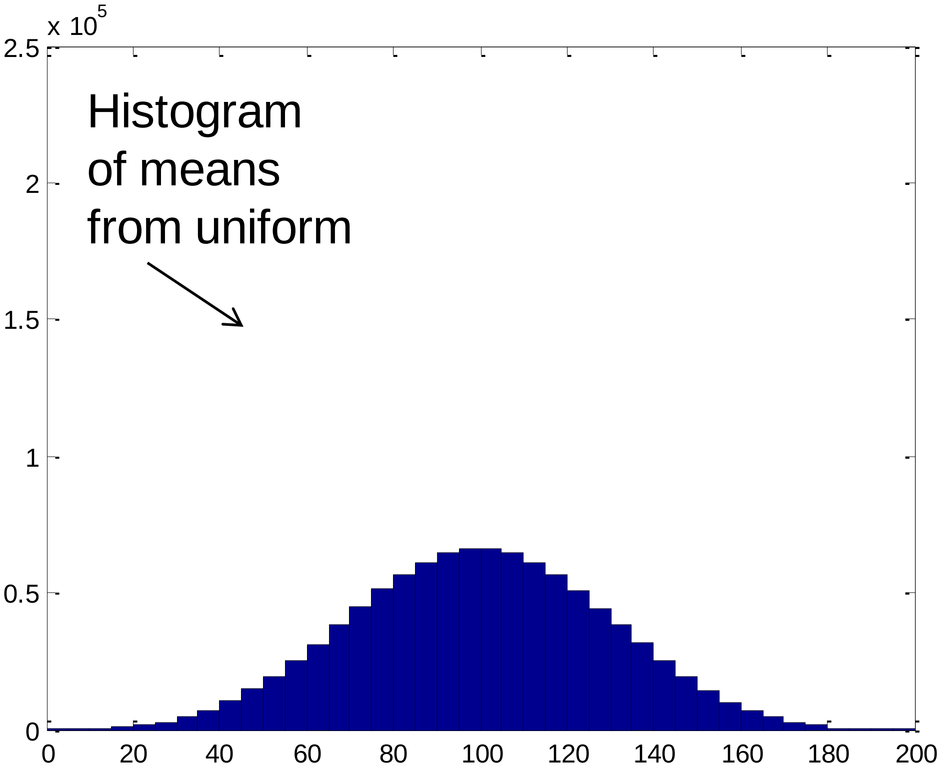


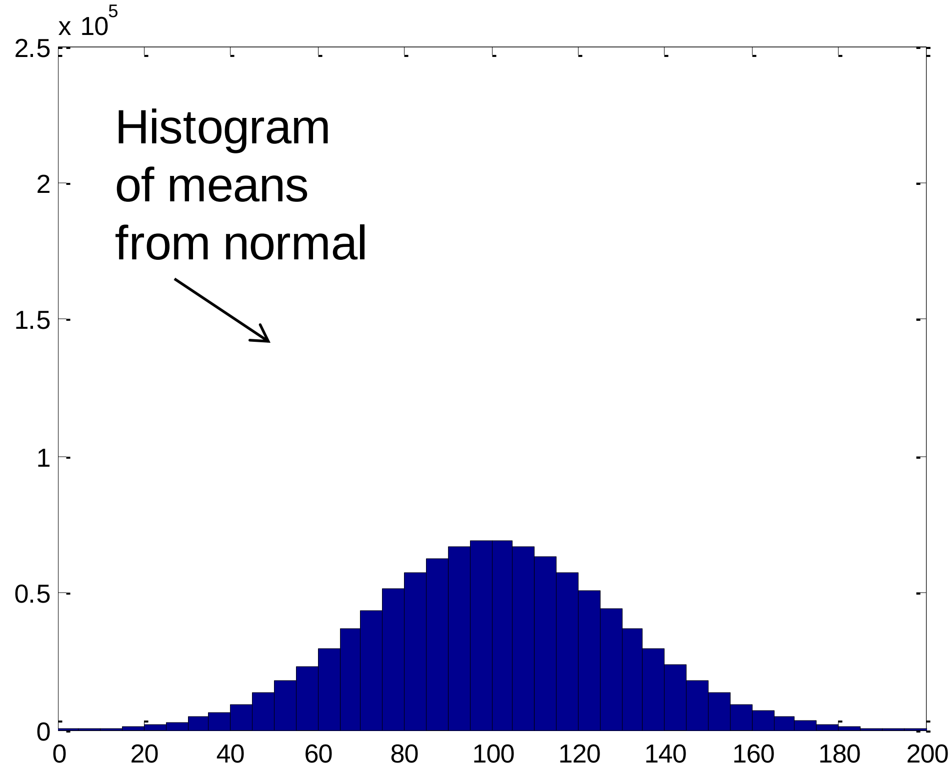


The Central Limit Theorem (CLT)

$$\mu = 100$$
$$\sigma = 57.73$$

n=4 1×10^6 means

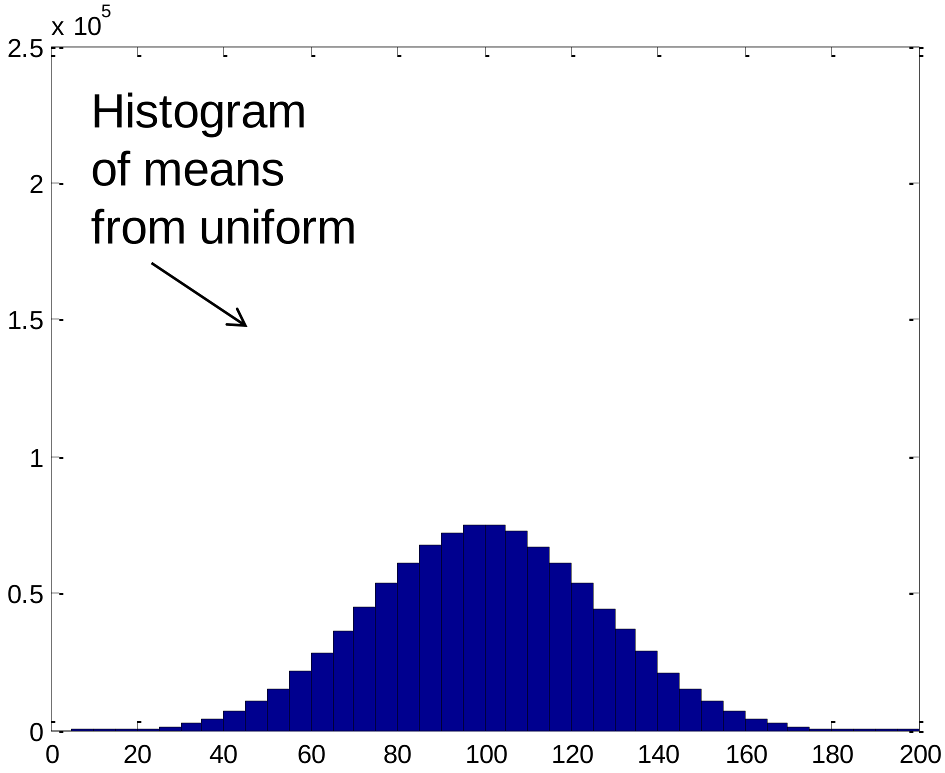


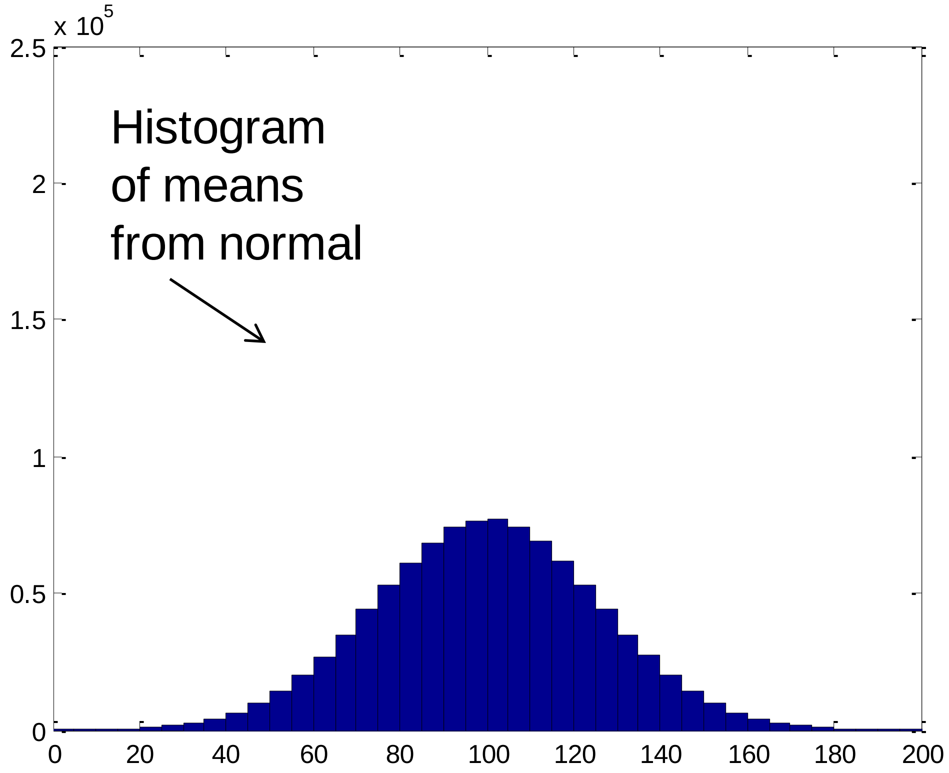


The Central Limit Theorem (CLT)

$\mu = 100$
$\sigma = 57.73$

n=5 1×10^6 means

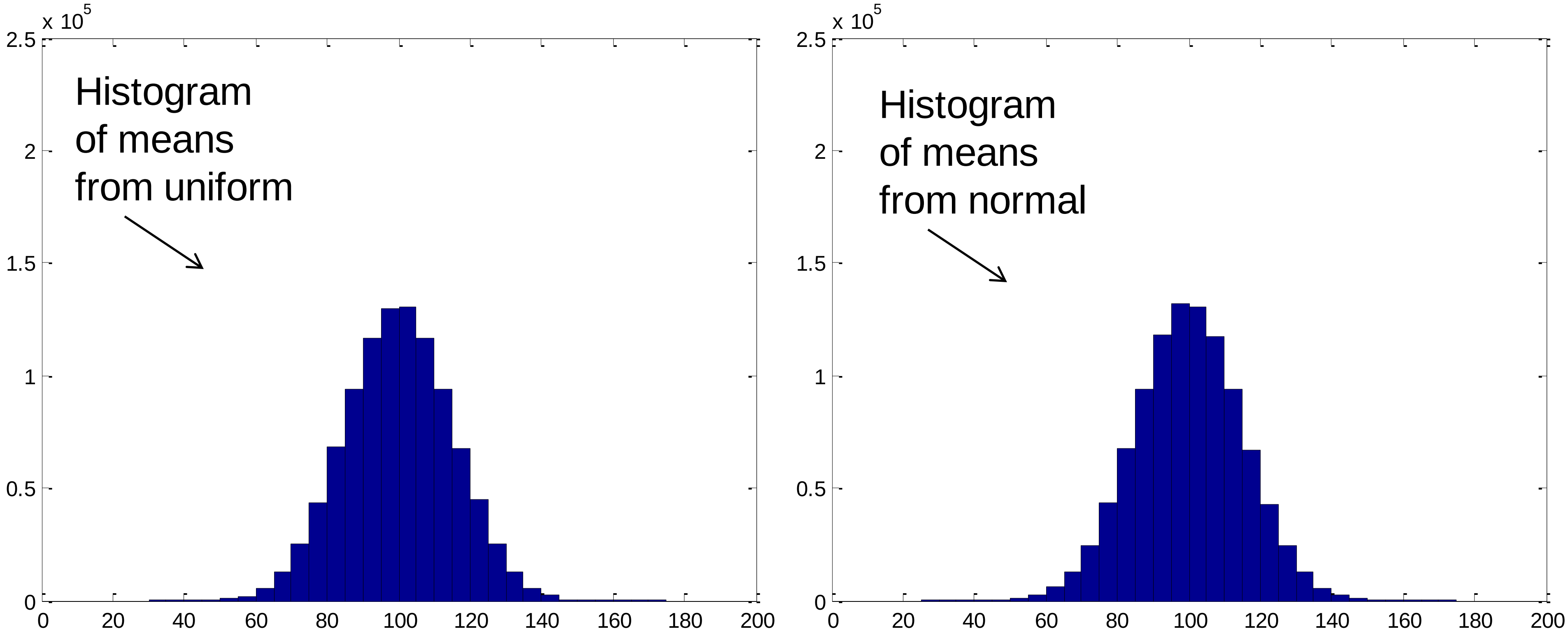




The Central Limit Theorem (CLT)

$$\mu = 100$$
$$\sigma = 57.73$$

n=15 1×10^6 means

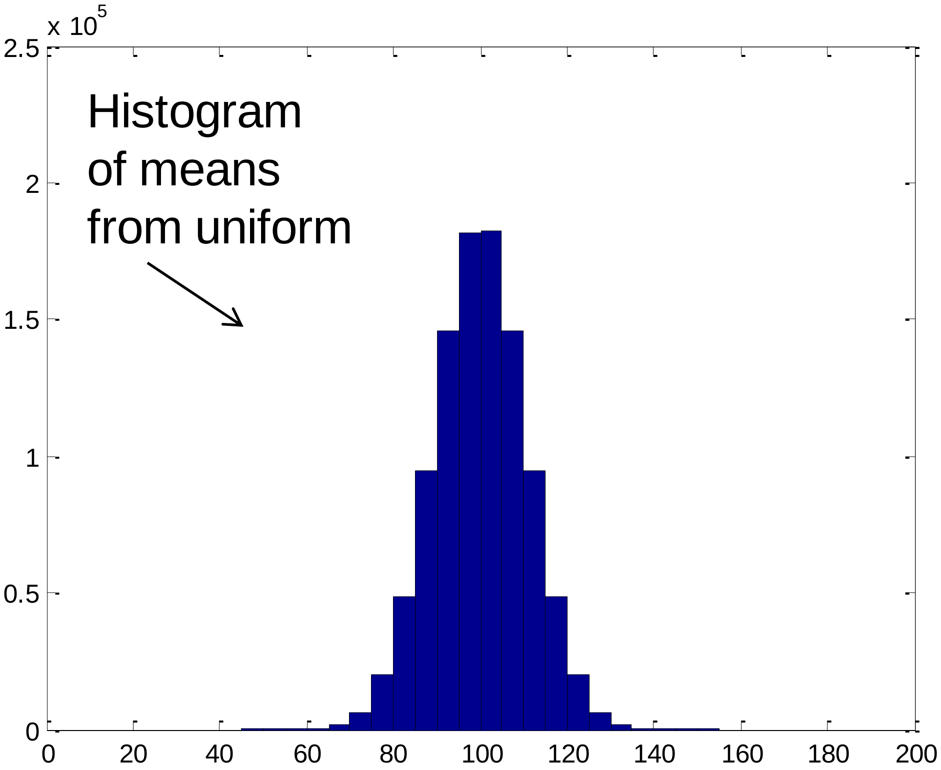


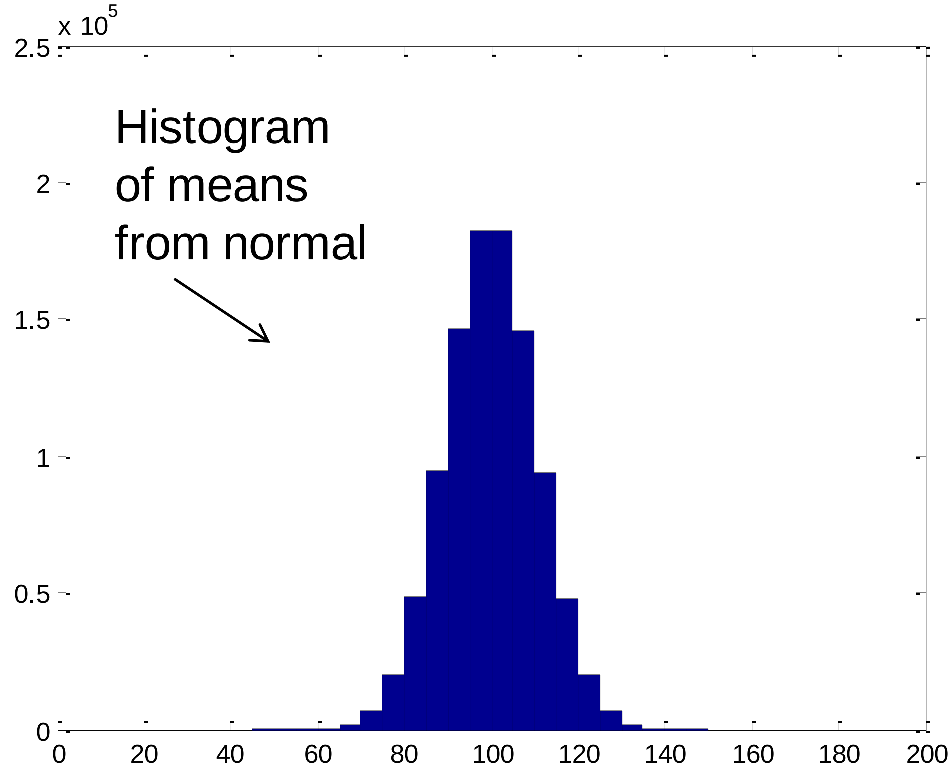
The Central Limit Theorem (CLT)

$\mu = 100$

$\sigma = 57.73$

n=30 $1{\times}10^6$ means



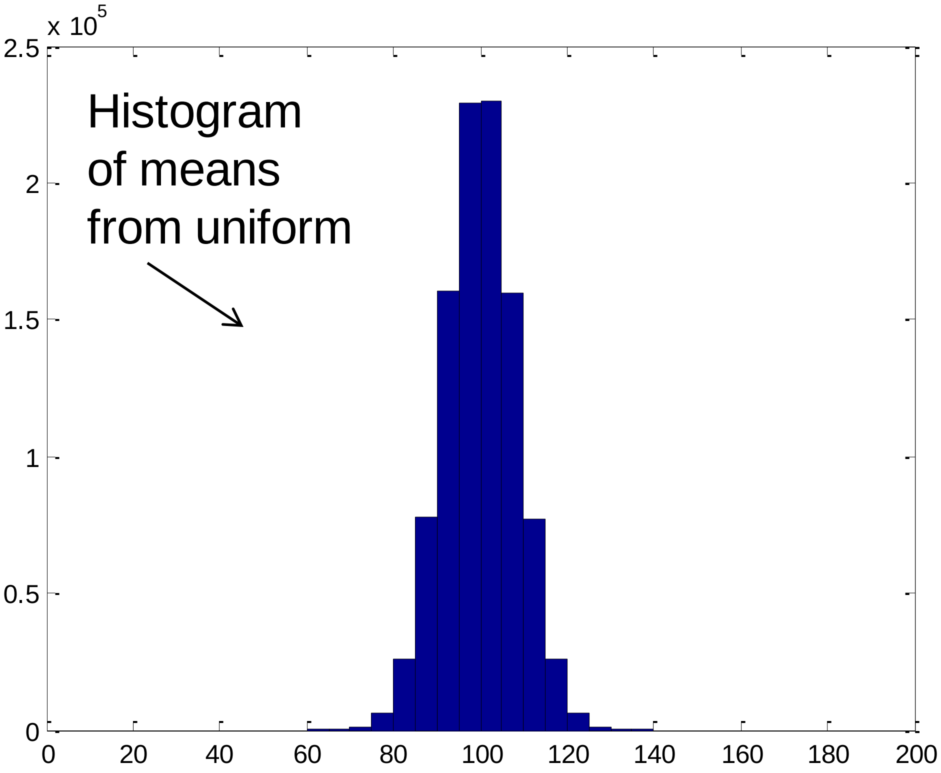


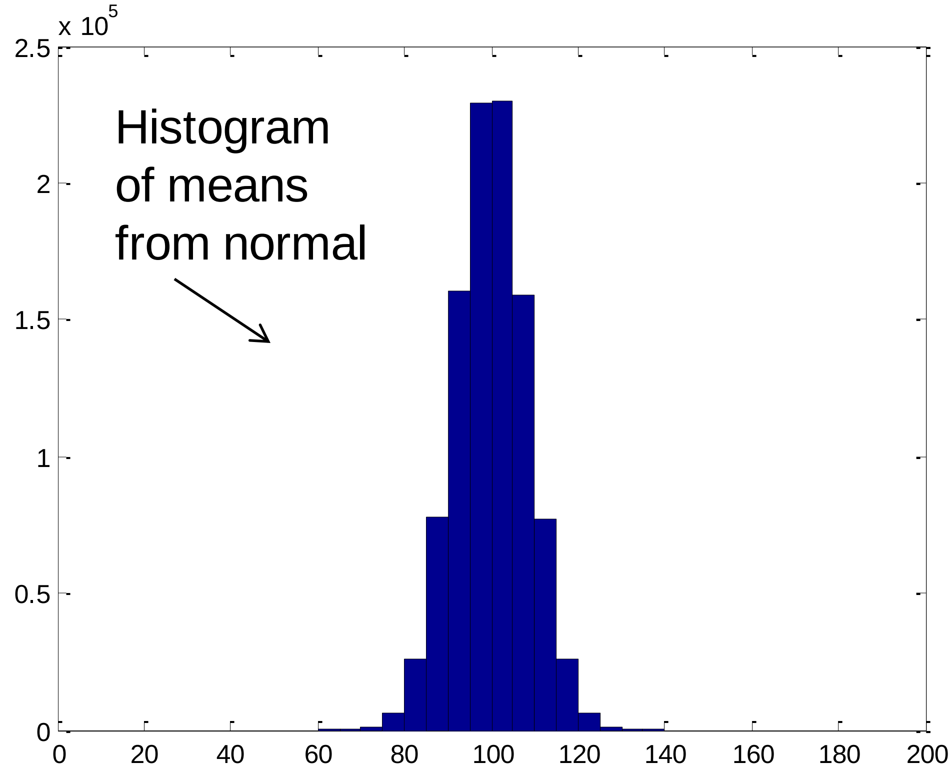
The Central Limit Theorem (CLT)

$\mu = 100$

$\sigma = 57.73$

n=50 1×10^6 means





Bivariate Transformation of Variables
(continued)

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

Outline

- Distributions

  Uniforms to Normals

  Normals to Chi-Square

  Normal and Chi-Square to t

  Chi-Squares to F

Bivariate Change of Variable - Normals

Let u_1~uniform(0,1) and u_2~uniform(0,1).
The joint PDF of (u_1,u_2) is

$$f(u_1,u_2)=\begin{cases}1 & \text{if} \quad u_1\in[0,1] \text{ and } u_2\in[0,1] \\ 0 & \text{if} \quad u_1\notin[0,1] \text{ or } u_2\notin[0,1]\end{cases}.$$

If $z_1=z_1(u_1,u_2)$, $z_2=z_2(u_1,u_2)$, the joint distribution of (z_1,z_2) is

$$f_{Z_1,Z_2}(z_1,z_2\,|\,\theta)=f_{U_1,U_2}\left(u_1(z_1,z_2),u_2(z_1,z_2)\,|\,\theta\right)\times|J(u_1,u_2\to z_1,z_2)|$$

$$J(u_1,u_2\to z_1,z_2)=\begin{vmatrix}\dfrac{du_1(z_1,z_2)}{dz_1} & \dfrac{du_1(z_1,z_2)}{dz_2} \\ \dfrac{du_2(z_1,z_2)}{dz_1} & \dfrac{du_2(z_1,z_2)}{dz_2}\end{vmatrix}$$

Bivariate Change of Variable - Normals

Let $z_1 = \sqrt{-2\ln(u_1)}\cos(2\pi u_2)$ and $z_2 = \sqrt{-2\ln(u_1)}\sin(2\pi u_2)$

then $u_1(z_1, z_2) = e^{-\frac{1}{2}(z_1^2 + z_2^2)}$ and $u_2(z_1, z_2) = \frac{1}{2\pi}\operatorname{atan}\left(\frac{z_2}{z_1}\right)$.

$$J(u_1, u_2 \to z_1, z_2) = \begin{vmatrix} \dfrac{du_1(z_1, z_2)}{dz_1} & \dfrac{du_1(z_1, z_2)}{dz_2} \\ \dfrac{du_2(z_1, z_2)}{dz_1} & \dfrac{du_2(z_1, z_2)}{dz_2} \end{vmatrix} = -\frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2 + z_2^2)}$$

Bivariate Change of Variable - Normals

Therefore,

$$f_{Z_1,Z_2}(z_1,z_2\,|\,\theta) = f_{U_1,U_2}\big(u_1(z_1,z_2),u_2(z_1,z_2)\,|\,\theta\big)\times |J(u_1,u_2\rightarrow z_1,z_2)|$$

which upon insertion yields

$$f_{Z_1,Z_2}(z_1,z_2\,|\,\theta) = \frac{1}{2\pi}e^{-\frac{1}{2}(z_1^2+z_2^2)}$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_2^2} \quad .$$

Joint PDF factors thus independent

This means $z_1 \sim N(0,1)$, $z_2 \sim N(0,1)$, z_1 and z_2 are independent.

Bivariate Change of Variable - Chi-Square

We discussed how we can obtain a random variable x

that has a general normal distribution with mean μ and

variance σ^2 via the transformation $x = \sigma z + \mu$.

The PDF of x can be obtained by the change of variable

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where, $x, \mu \in \mathbb{R}$, $0 < \sigma$. That is, x~normal(μ,σ^2).

Bivariate Change of Variable - Chi-Square

We also discussed how the change of variable technique can be repeated. If x_i~normal(μ,σ^2) for i=1,…,n, and x_i's are independent, then

$$y = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad .$$

Bivariate Change of Variable - Chi-Square

We also discussed how the change of variable technique can be applied to $y_1 = \left(\frac{x_1 - \mu}{\sigma} \right)^2$. If x_1~normal(μ,σ^2), then the distribution y_1 is $\chi^2(1)$. This process can be duplicated so that if x_2~normal(μ,σ^2), then the distribution of

$$y_2 = \left(\frac{x_2 - \mu}{\sigma} \right)^2 \quad \text{is } \chi^2(1).$$

Now what is the distribution of $y_1 + y_2$?

Bivariate Change of Variable - Chi-Square

Let y_1 and y_2 have independent chi-square PDFs

$$f(y_i)=\frac{y_i^{1/2-1}e^{-y_i/2}}{\Gamma(1/2)2^{1/2}}\,, \quad y_i>0,\ i=1,2.$$

We can find the distribution of $w_1=y_1+y_2$ (and w_2=y_2)

via the bivariate change of variable technique

$$f_{W_1,W_2}(w_1,w_2\,|\,\theta)=f_{Y_1,Y_2}(y_1(w_1,w_2),y_2(w_1,w_2)\,|\,\theta)\times|J(y_1,y_2\to w_1,w_2)|$$

with marginalization $f_{W_1}(w_1\,|\,\theta)=\int_{w_2} f_{W_1,W_2}(w_1,w_2\,|\,\theta)\,dw_2.$

Bivariate Change of Variable - Chi-Square

It turns out that if $y_1 \sim \chi^2(1)$, $y_2 \sim \chi^2(1)$, and independent, then

$w_1 = y_1 + y_2 \sim \chi^2(2)$. Or more generally, if $y_1 \sim \chi^2(\nu_1)$,

$y_2 \sim \chi^2(\nu_2)$, and independent, then $w_1 = y_1 + y_2 \sim \chi^2(\nu_1 + \nu_2)$.

So what this means is that

$$y = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)\,!$$

⟵ Homework problem.

Bivariate Change of Variable - Chi-Square

If the mean μ is unknown, then we can estimate it by $\bar{x}$ and lose one degree of freedom!

$$\underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}_{\chi(n)} = \underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}_{\chi(n-1)} + \underbrace{\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}_{\chi(1)}$$

We just showed

add and subtract $\bar{x}$ in the numerator

Because df add, or by transformation!

Since $\bar{x} \sim N(\mu,\sigma^2/n)$

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$

Bivariate Change of Variable - Student-t

We showed that if x_i~normal(μ,σ^2) for i=1,…,n, then

the distribution of $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

and that the distribution of $y_2 = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \sim \chi(n-1)$.

Note that $y_2 = \frac{(n-1)s^2}{\sigma^2}$.

It turns out that z and $\frac{(n-1)s^2}{\sigma^2}$ are statistically independent!

Bivariate Change of Variable - Student-t

So $z=\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ and $y_2=\dfrac{\nu s^2}{\sigma^2}\sim\chi(\nu)$, $\nu=n-1$.

Let $t=\dfrac{z}{\sqrt{y_2/\nu}}$ and $s=y_2$.

Then $z=\dfrac{t\sqrt{s}}{\sqrt{\nu}}$ and $y_2=s$, the Jacobian of the transformation is

$$J(z,y\to t,s)=\begin{vmatrix}\dfrac{dz(t,s)}{dt} & \dfrac{dz(t,s)}{ds}\\ \dfrac{dy_2(t,s)}{dt} & \dfrac{dy_2(t,s)}{ds}\end{vmatrix}=\frac{\sqrt{s}}{\sqrt{\nu}}$$

Bivariate Change of Variable - Student-t

Here we use the assumption that z and y are independent!

The joint distribution of (t,s) is

$$f_{T,S}(t,s|\theta) = f_{y_2,z}(y_2(t,s),z(t,s)|\theta)\times|J(y_2,z\to t,s)|$$

$$f_{T,S}(t,s|\theta) = \frac{s^{\frac{\nu}{2}-1}e^{-\frac{\nu}{2}\left(1+\frac{1}{\nu}t^2\right)}}{\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}\sqrt{2\pi}}\times\left|\frac{\sqrt{s}}{\sqrt{\nu}}\right|$$

and by integrating out s the distribution of t is

$$f_T(t|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{1}{\nu}t^2\right)^{-\frac{\nu+1}{2}}.$$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$y_2 = \frac{(n-1)s^2}{\sigma^2}$$

The distribution of $t = \dfrac{z}{\sqrt{y_2/(n-1)}} \sim t(n-1)$!

Bivariate Change of Variable - F

Recall that
$$\underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}_{\chi(n)}=\underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}_{\chi(n-1)}+\underbrace{\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}_{\chi(1)}\quad,$$

It turns out that $y_1=\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2$ and $y_2=\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2$

are statistically independent.

But of interest to us (hypothesis testing) is the distribution of

$f=\dfrac{y_1/\nu_1}{y_2/\nu_2}$, where $y_1\sim\chi^2(\nu_1)$ and $y_2\sim\chi^2(\nu_2)$.

Bivariate Change of Variable - F

Let y_1 and y_2 have independent χ^2 PDFs with ν_1 and ν_2 df

$$f(y_i \mid \nu_i) = \frac{y_i^{\nu_i/2-1} e^{-y_i/2}}{\Gamma(\nu_i/2) 2^{\nu_i/2}} \ , \quad y_i > 0 \, , \ i = 1,2 \ .$$

We can find the distribution of $f = \dfrac{y_1/\nu_1}{y_2/\nu_2}$ (and g=y_2)

via the bivariate change of variable technique

$$f_{F,G}(f,g \mid \theta) = f_{Y_1,Y_2}(y_1(f,g), y_2(f,g) \mid \theta) \times |J(y_1, y_2 \to f,g)|$$

and marginalization $f_F(f \mid \theta) = \int_g f_{F,G}(f,g \mid \theta) dg$.

Bivariate Change of Variable - F

The joint distribution of (f,g) is

$$f_{F,G}(f,g|\theta)=f_{Y_1,Y_2}(y_1(f,g),y_2(f,g)|\theta)\times|J(y_1,y_2\to f,g)|$$

the original variables in terms of the new variables are

$y_1=\frac{\nu_1}{\nu_2}gf$ and $y_2=g$ with Jacobian

$$J(y_1,y_2\to f,g)=\begin{vmatrix}\frac{dy_1(f,g)}{df} & \frac{dy_1(f,g)}{dg}\\ \frac{dy_2(f,g)}{df} & \frac{dy_2(f,g)}{dg}\end{vmatrix}=\frac{\nu_1}{\nu_2}g \quad .$$

Bivariate Change of Variable - F

$$y_1 = \frac{\nu_1}{\nu_2} gf \qquad y_2 = g$$

The joint distribution of (f, g) is

$$f_{F,G}(f,g|\theta) = f_{Y_1,Y_2}(y_1(f,g), y_2(f,g)|\theta) \times |J(y_1,y_2 \to f,g)|$$

$$f_{F,G}(f,g|\theta) = \frac{\left(\frac{\nu_1}{\nu_2} gf\right)^{\nu_1/2-1} e^{-\left(\frac{\nu_1}{\nu_2} gf\right)/2}}{\Gamma(\nu_1/2)2^{\nu_1/2}} \frac{g^{\nu_2/2-1} e^{-g/2}}{\Gamma(\nu_2/2)2^{\nu_2/2}} \times \left|\frac{\nu_1}{\nu_2} g\right|$$

$$f_F(f|\theta) = \int_g f_{F,G}(f,g|\theta) dg$$

$$f_F(f|\nu_1,\nu_2) = \frac{\Gamma((\nu_1+\nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1 f}{\nu_1 f + \nu_2}\right)^{\nu_1/2} \left(1 - \frac{\nu_1 f}{\nu_1 f + \nu_2}\right)^{\nu_2/2}$$

Bivariate Change of Variable - F

The joint distribution of $f = \dfrac{y_1 / \nu_1}{y_2 / \nu_2}$ is

F distributed with ν_1 numerator df and ν_2 denominator df

$$f_F(f \mid \nu_1, \nu_2) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1 f}{\nu_1 f + \nu_2}\right)^{\nu_1/2} \left(1 - \frac{\nu_1 f}{\nu_1 f + \nu_2}\right)^{\nu_2/2}$$

where $\nu_1, \nu_2 = 1, 2, \ldots$

$$\underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}_{\chi(n)} = \underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}_{\chi(n-1)} + \underbrace{\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}_{\chi(1)}$$

Therefore,

$$F = \left[\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 \Big/ 1\right] \Big/ \left[\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \Big/ (n-1)\right] \sim F(1, n-1)$$

Bivariate Change of Variable - F/Student-t

We just showed that

$$f = \frac{y_1 / \nu_1}{y_2 / \nu_2} \sim F(1, n-1) \text{ where } y_1 = \left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right)^2 \text{ and } y_2 = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^2$$

Recall that we showed that

$$t = \frac{z}{\sqrt{y_2/(n-1)}} \sim t(n-1) \text{ where } z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \text{ and } y_2 = \frac{(n-1)s^2}{\sigma^2} ?$$

What this means is, when $\nu_1 = 1$, $f = t^2$!

$$t^2 = F = \left[\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right)^2 \Big/ 1 \right] \Bigg/ \left[\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \Big/ (n-1) \right]$$

Bivariate Change of Variable - normal, χ^2, t, F

Recap: u_1 and u_2~uniform(0,1) and independent

$$z_1 = \sqrt{-2\ln(u_1)}\cos(2\pi u_2) \qquad z_2 = \sqrt{-2\ln(u_1)}\sin(2\pi u_2)$$

$z_1 \sim N(0,1)$, $z_2 \sim N(0,1)$, z_1 and z_2 are independent

$$x_i = \sigma z_i + \mu \sim N(\mu,\sigma^2), \quad \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$y_1 = \left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1), \quad y_2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

y_1 and y are independent

$$t = \frac{z}{\sqrt{y_2/(n-1)}} \sim t(n-1), \qquad f = \frac{y_1/1}{y_2/(n-1)} \sim F(1,n-1).$$

Multivariate Transformation of Variables and Maximum Likelihood Estimation

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

MARQUETTE UNIVERSITY
Be The Difference.

Outline

- **Multivariate Transformation of Variables**

- **Maximum Likelihood Estimation (MLE)**

Multivariate Change of Variable

The distribution of $(y_1,\ldots,y_n)$ is

$$f_Y(y_1,\ldots,y_n\,|\,\theta) = f_X\left(x_1(y_1,\ldots,y_n),\ldots,x_n(y_1,\ldots,y_n)\,|\,\theta\right) \times |J(x_1,\ldots,x_n \to y_1,\ldots,y_n)|$$

$$f_Y(y_1,\ldots,y_n) = \left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[(y_1-\sum_{i=2}^{n}y_i-\mu)^2+\sum_{i=2}^{n}(y_i+y_1-\mu)^2\right]\right\}|n|$$

WOLOG let $\mu=0, \sigma=1$

$$\bar{x} \swarrow$$

$$f_Y(y_1,\ldots,y_n) = \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}y_1^2}\cdot\frac{n^{-\frac{1}{2}}}{(2\pi)^{(n-1)/2}}e^{-\frac{1}{2}\left[\sum_{i=2}^{n}y_i^2+\left(\sum_{i=2}^{n}y_i\right)^2\right]}$$

$$\nearrow \quad s^2 = \frac{1}{n-1}\left(\left[\sum_{i=2}^{n}(x_i-\bar{x})\right]^2+\sum_{i=2}^{n}(x_i-\bar{x})^2\right)$$

So $\bar{x}$ and s^2 are independent.

Multivariate Change of Variable

The important moral to learn from our study of

transformation of variables is:

Measurements have statistical variation and a statistical

distribution associated with them and every time we do

something with a measurement (i.e. math operation on it)

we change its statistical properties and its distribution!

Maximum Likelihood Estimation - Mean

$L(\mu,\sigma^2)$ is called the likelihood function.

What we want to do is find the values of (μ,σ^2)

that maximize $L(\mu,\sigma^2)$. The value of μ that maximizes

$L(\mu,\sigma^2)$ is the value $\hat{\mu}$ that minimizes $\sum_{i=1}^{n}(y_i-\mu)^2$.

The value of σ^2 that maximizes $L(\mu,\sigma^2)$ is

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu})^2 .$$

$d_i = y_i - \hat{\mu}$

minimize $\sum_{i=1}^{n} d_i^2$

Maximum Likelihood Estimation - Mean

$L(\mu,\sigma^2)$ is called the likelihood function.

What we do is differentiate $L(\mu,\sigma^2)$ wrt μ and σ^2, set = 0 and solve. That is,

$$\left.\frac{\partial L(\mu,\sigma^2)}{\partial \mu}\right|_{\hat{\mu},\hat{\sigma}^2}=0 \quad \text{and} \quad \left.\frac{\partial L(\mu,\sigma^2)}{\partial \sigma^2}\right|_{\hat{\mu},\hat{\sigma}^2}=0 .$$

However, this is messy, but we can instead maximize

$$LL(\mu,\sigma^2)=\ln(L(\mu,\sigma^2))$$

because it is a monotonic function. Use $\log(\cdot)$ for $\ln(\cdot)$.

Maximum Likelihood Estimation - Mean

With $y_i = \mu + \varepsilon_i$ and $\varepsilon_i \sim N(0,\sigma^2)$, ε_i independent,

$$f(y_1,\ldots,y_n \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

$$LL(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

$$\left.\frac{\partial LL(\mu,\sigma^2)}{\partial\mu}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i-\hat{\mu})(-1) = 0$$

$$\left.\frac{\partial LL(\mu,\sigma^2)}{\partial\sigma^2}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{n}{2}\frac{2}{\hat{\sigma}^2} - \frac{-1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(y_i-\hat{\mu})^2 = 0$$

Maximum Likelihood Estimation - Mean

Solving for μ and σ^2 yields

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2 \ .$$

These are MLEs, most probable or modal values.

Note that the denominator is n and not n-1.

This is why we use a denominator n-1.

$\hat{\sigma}^2$ is a biased estimator of σ^2, $E(\hat{\sigma}^2) = \frac{(n-1)}{n}\sigma^2$.

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1) \longrightarrow E\left(\frac{n\hat{\sigma}^2}{\sigma^2}\right) = n-1 \longrightarrow E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

Maximum Likelihood Estimation - Linear

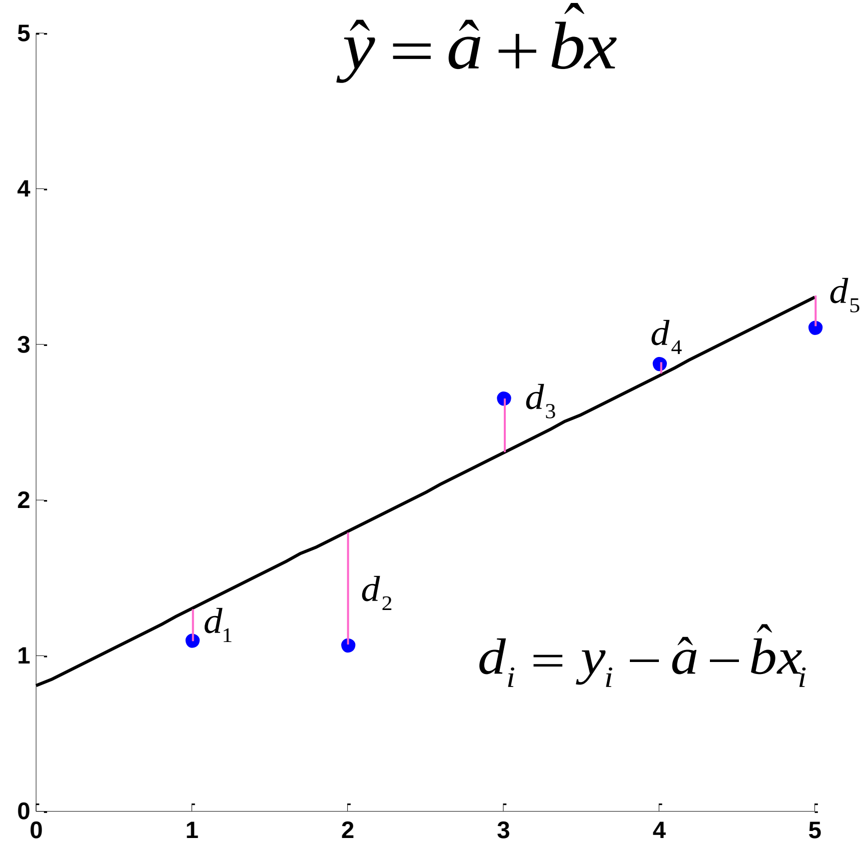
This technique, can be generalized to linear regression.

Let $y_i = a + bx_i + \varepsilon_i$,

where $\varepsilon_i \sim N(0,\sigma^2)$

are independent.

$i = 1,\ldots,n$



Maximum Likelihood Estimation - Linear

This technique, can be generalized to linear regression.

Let $y_i = a + bx_i + \varepsilon_i$, where $\varepsilon_i \sim N(0,\sigma^2)$ are independent.

Then, the likelihood is

$$f(y_1,\ldots,y_n \mid a,b,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2\right]$$

and the log likelihood is

$$LL(a,b,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2 .$$

Maximum Likelihood Estimation - Linear

Differentiate $LL(a,b,\sigma^2)$ wrt a, b, and σ[2], then set = 0

$$LL(a,b,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial a}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i - \hat{a} - \hat{b}x_i)(-1) = 0$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial b}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i - \hat{a} - \hat{b}x_i)(-x_i) = 0$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial \sigma^2}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{n}{2}\frac{2}{\hat{\sigma}^2} - \frac{-1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)^2 = 0$$

Maximum Likelihood Estimation - Linear

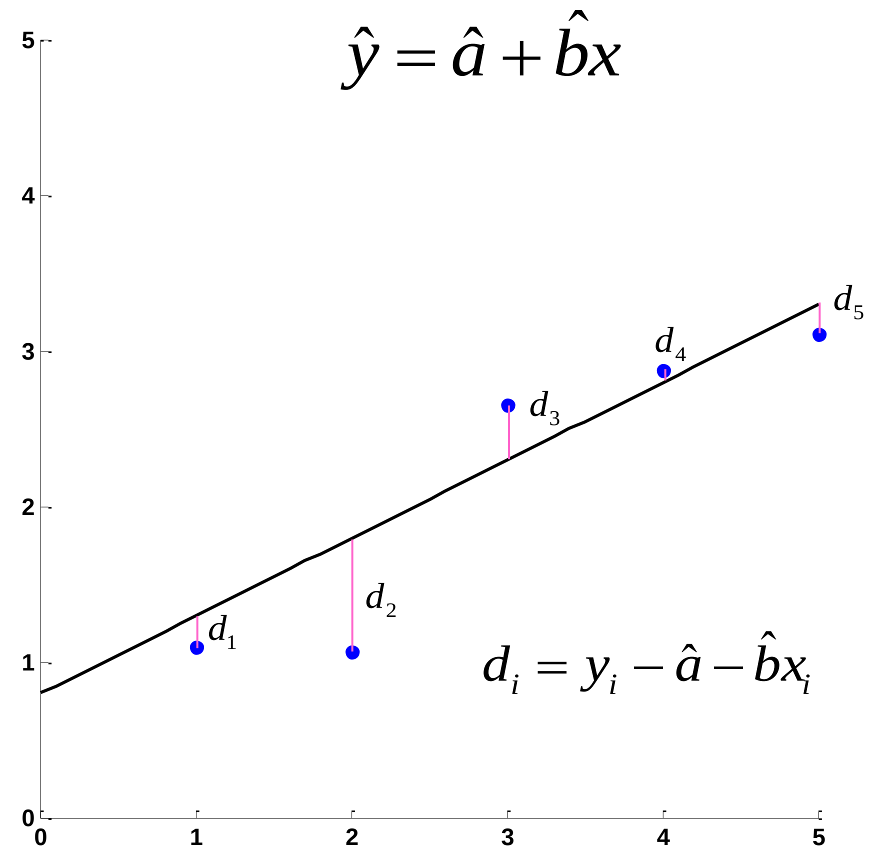
Solving for the estimated parameters yields

$$\hat{b} = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{a} = \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2$$



Maximum Likelihood Estimation - Linear

The regression model $y_i = a + bx_i + \varepsilon_i$ where $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$
$i = 1,\ldots,n$

that we presented, can be equivalently written as

$y = X\beta + \varepsilon$ where

measured data, design matrix, regression coefficients, measurement error

$$\underset{n\times 1}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \underset{n\times 2}{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \underset{2\times 1}{\beta} = \begin{pmatrix} a \\ b \end{pmatrix}, \quad \underset{n\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

and $\varepsilon \sim N(0,\sigma^2 I_n)$. I_n is an n-dimensional identity matrix.

Maximum Likelihood Estimation - Linear

With $\underset{n\times 1}{y} = X\beta + \varepsilon$ and $\underset{n\times 1}{\varepsilon} \sim N(0,\sigma^2 I_n)$

The likelihood is

$$f(y_1,\ldots,y_n \mid a,b,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]$$

and the log likelihood is

$$LL(a,b,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

Maximum Likelihood Estimation - Linear

$L(\beta,\sigma^2)$ is again called the likelihood function.

What we want to do is find the values of (β,σ^2)

that maximize $L(\beta,\sigma^2)$. The value of β that maximizes

$L(\beta,\sigma^2)$ is the value $\hat{\beta}$ that minimizes $(y-X\beta)'(y-X\beta)$.

The value of σ^2 that maximizes $L(\beta,\sigma^2)$ is

$$\hat{\sigma}^2=\frac{1}{n}(y-X\hat{\beta})'(y-X\hat{\beta})$$

We need to find $\hat{\beta}$.

$d_i = y_i - \hat{a} - \hat{b}x_i$

minimize $(y-X\beta)'(y-X\beta)$ wrt β

Maximum Likelihood Estimation - Linear

We don't need to take the derivative of $L(\beta,\sigma^2)$

wrt β (although we could). We can write with algebra

$$(y-X\beta)'(y-X\beta)=(y-X\hat{\beta})'(y-X\hat{\beta})+(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})$$

↖ add and subtract $X\beta$

where $\hat{\beta}=(X'X)^{-1}X'y$, it can be seen that $\beta=\hat{\beta}$ maximizes

↖ invertible

$$LL(\beta,\sigma^2)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)$$

$$-\frac{1}{2\sigma^2}\left[(y-X\hat{\beta})'(y-X\hat{\beta})+(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})\right]$$

Maximum Likelihood Estimation - Linear

More generally, we can have a multiple regression model

$$\underset{n\times 1}{y} = X\beta + \varepsilon \text{ where } \underset{n\times 1}{\varepsilon} \sim N(0,\sigma^2 I_n) \text{ and}$$

measured data — design matrix — regression coefficients — measurement error

$$\underset{n\times 1}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \underset{n\times(q+1)}{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1q} \\ 1 & x_{21} & & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nq} \end{pmatrix}, \quad \underset{(q+1)\times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}, \quad \underset{n\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Maximum Likelihood Estimation - Linear

The MLEs are the same,

$$\underset{(q+1)\times 1}{\hat{\beta}} = (X'X)^{-1}X'y \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}) .$$

In addition,

$$\underset{(q+1)\times 1}{\hat{\beta}} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right) \quad \text{and} \quad n\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-q-1) .$$

$$\underbrace{(y - X\beta)'(y - X\beta)}_{\chi^2(n)} = \underbrace{(y - X\hat{\beta})'(y - X\hat{\beta})}_{\chi^2(n-q-1)} + \underbrace{(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})}_{\chi^2(q+1)}$$

This means we should use a denominator of n-q-1 for unbiased estimator of σ^2.

independent

Maximum Likelihood Estimation - Exponential

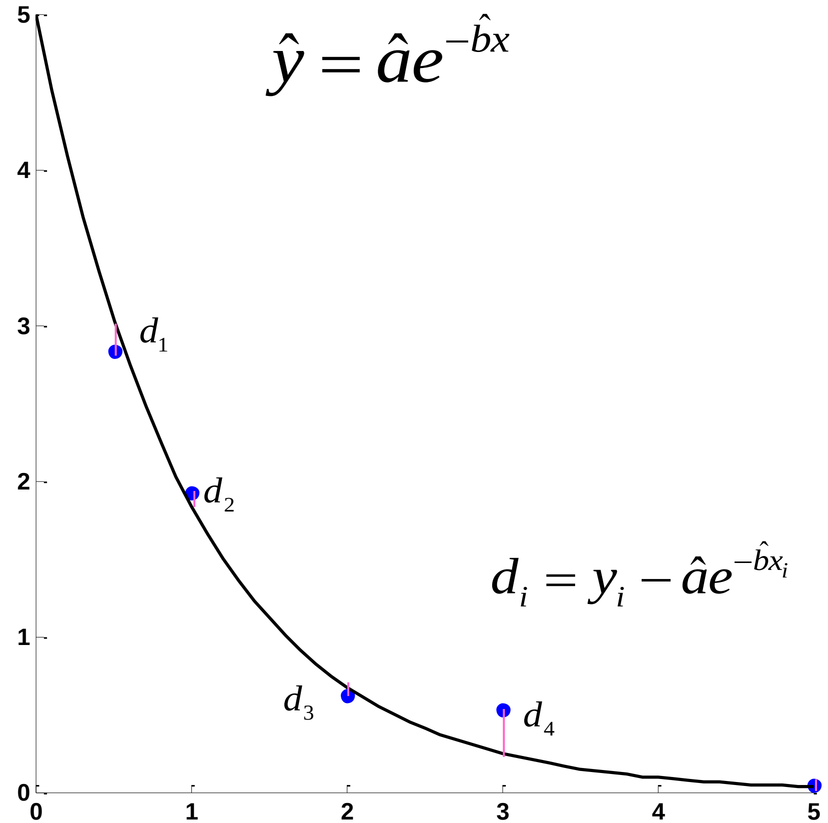
This is a more general method than just for linear functions

Let $y_i = ae^{-bx_i} + \varepsilon_i$,

where $\varepsilon_i \sim N(0,\sigma^2)$

are independent.

$i = 1,\ldots,n$



Maximum Likelihood Estimation - Exponential

This is a more general method than just for linear functions

Let $y_i = ae^{-bx_i} + \varepsilon_i$, where $\varepsilon_i \sim N(0,\sigma^2)$ are independent.

Then, the likelihood is

$$f(y_1,\ldots,y_n \mid a,b,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - ae^{-bx_i})^2\right]$$

and the log likelihood is

$$LL(a,b,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - ae^{-bx_i})^2 .$$

Maximum Likelihood Estimation - Exponential

$L(a,b,\sigma^2)$ is again called the likelihood function.

What we want to do is find the values of (a,b,σ^2)

that maximize $L(a,b,\sigma^2)$. The values (a,b) that maximize

$L(a,b,\sigma^2)$ are the values $(\hat{a},\hat{b})$ that minimize $\sum_{i=1}^{n}(y_i - ae^{-bx_i})^2$.

The value of σ^2 that maximizes $L(a,b,\sigma^2)$ is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{a}e^{-\hat{b}x_i})^2 .$$

$d_i = y_i - \hat{a}e^{-\hat{b}x_i}$

minimize $\sum_{i=1}^{n} d_i^2$
wrt a, b

Maximum Likelihood Estimation - Exponential

Differentiate $LL(a,b,\sigma^2)$ wrt a, b, and σ², then set = 0

$$LL(a,b,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - ae^{-bx_i})^2$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial a}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i - \hat{a}e^{-\hat{b}x_i})(-e^{-\hat{b}x_i}) = 0$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial b}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i - \hat{a}e^{-\hat{b}x_i})(-x_i e^{-\hat{b}x_i}) = 0$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial \sigma^2}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{n}{2}\frac{2}{\hat{\sigma}^2} - \frac{-1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(y_i - \hat{a}e^{-\hat{b}x_i})^2 = 0$$

Maximum Likelihood Estimation - Exponential

Solving for the estimated parameters yields

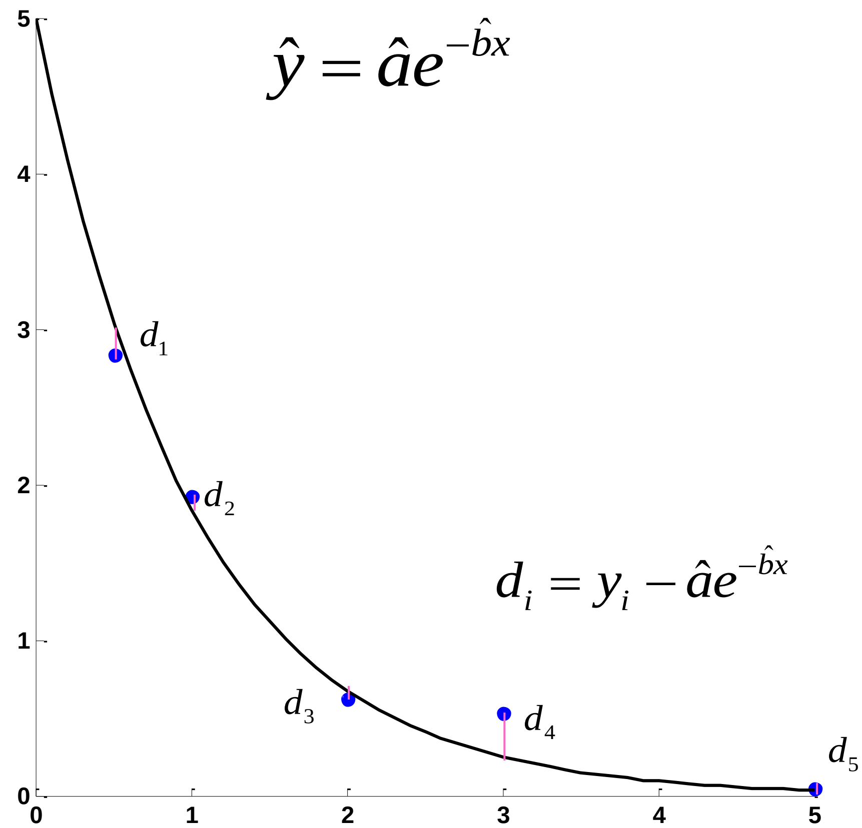
$$\hat{a} = \frac{\sum_{i=1}^{n} y_i e^{-\hat{b}x_i}}{\sum_{i=1}^{n} e^{-\hat{b}x_i}}$$

No analytic solution.

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i e^{-\hat{b}x_i}}{\sum_{i=1}^{n} x_i e^{-2\hat{b}x_i}}$$

Need numerical Solution.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{a}e^{-\hat{b}x_i})^2$$



Maximum Likelihood Estimation - Exponential

Since we had to numerically maximize the likelihood,

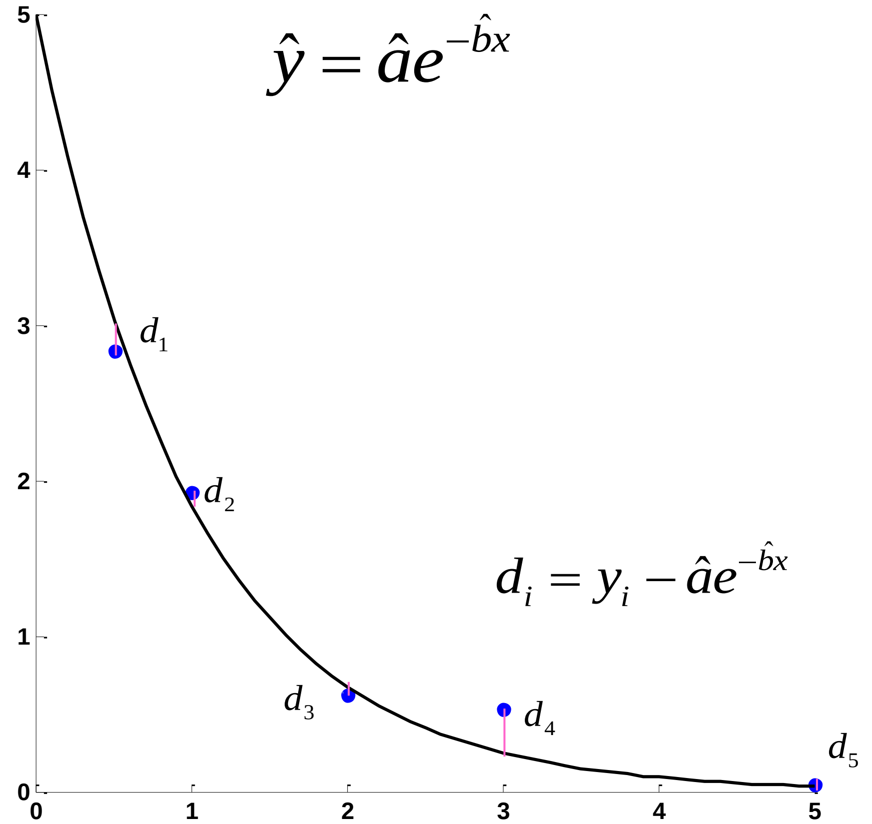
we do not have "nice" formulas

for the mean and variance of

(a,b,σ^2)

a and b that minimize $\sum_{i=1}^{n} d_i^2$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{a}e^{-\hat{b}x_i})^2$$



Hypothesis Testing

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

MARQUETTE UNIVERSITY
Be The Difference.

Outline

- **Hypothesis Testing**

- **Likelihood Ratio Test (LRT) Background**

  Unconstrained and Constrained Maximization

- **LRT Examples**

  Mean, Linear Regression, Difference in 2 Means, Analysis of Variance (ANOVA)

Hypothesis Testing

Steps in Hypothesis testing.

Step 1: Set up hypotheses (state H_0 and H_1). Select α.

Step 2: Select appropriate test statistic.

Step 3: Generate decision rule.

Step 4: Compute test statistic.

Step 5: Draw a conclusion about H_0 by comparing test statistic in **Step 4** to decision rule in **Step 3**.

Hypothesis Testing

Example: Population mean. $y_i = \mu + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$
$i = 1,\ldots,n$

Step 1: Hypotheses. H_0: $\mu=\mu_0$, H_1: $\mu\neq\mu_0$. $\alpha=0.05$

Step 2: Test statistic. $t = \dfrac{\bar{y}-\mu_0}{s/\sqrt{n}}$ i.e. $\mu_0 = 100$, $n = 5$

Step 3: Generate decision rule. Reject H_0 if $|t| > t_{\alpha/2}(n-1)$.

Step 4: Compute test statistic. $t=3.14$ i.e. $t_{.05/2}(4) = 2.776$

Step 5: Draw a conclusion. Since $3.14 > 2.776$, reject H_0.

Likelihood Ratio Test - Background

The LRT procedure involves: $y_1,\ldots,y_n$ iid from $f_Y(y|\theta)$

1) Writing the likelihood function of the observations.

$$L(\theta)=\prod_{i=1}^{n} f_Y(y_i|\theta) \qquad \theta=(\theta,\ldots,\theta_p)$$

and this can be done instead with the log likelihood.

$$LL(\theta)=\sum_{i=1}^{n}\log(f_Y(y_i|\theta))$$

⟵ Recall we discussed max of LL(θ) instead of L(θ).

Likelihood Ratio Test - Background

2) Maximizing the log likelihood wrt θ assuming H_1 is true.

H_1: $g(\theta) \neq \theta_0$

max wrt θ $L(\theta)$ subject to constraint that H_1 true $\longrightarrow$ $\hat{\theta} = \hat{\theta}(y_1, \ldots, y_n)$

Since θ has no constraints it can be any value (except θ_0).

The log likelihood $LL(\theta \mid H_1) = LL(\theta)$

We actually fudge a little and maximize with θ_0 a possibility.

This may require numerical maximization.

Likelihood Ratio Test - Background

3) Maximizing the log likelihood wrt θ assuming H_0 is true.

H_0: $g(\theta) = \theta_0$

$$\max_{\text{wrt } \theta} L(\theta) \text{ subject to constraint that } H_0 \text{ true} \longrightarrow \tilde{\theta} = \tilde{\theta}(y_1, \ldots, y_n)$$

Since θ has constraints on it, we incorporate into $LL(\theta)$

The log likelihood $LL(\theta \mid H_0) = LL(\theta) + \delta(\underbrace{g(\theta) - \theta_0}_{\text{Constraint}})$

where δ is the Lagrange Multiplier.

This may require numerical maximization.

Likelihood Ratio Test - Background

4) Inserting values back in likelihoods and taking the ratio.

$\lambda = \dfrac{L(\tilde{\theta})}{L(\hat{\theta})}$, note $0 \le \lambda \le 1$.

Since λ is a function of the data $\tilde{\theta}(y_1,\ldots,y_n)$, $\hat{\theta}(y_1,\ldots,y_n)$
it is a statistic and has a distribution.

The interpretation is: $\longrightarrow$

λ =1 H_0 is true.

λ =0 H_1 is true.

Need to find cutoff c so that

$c \le \lambda \le 1$ Do not reject H_0

$0 \le \lambda < c$ Reject H_0.

Likelihood Ratio Test - Background

5) Test statistic and its distribution.

In large samples (and under mild regularity conditions)

$$-2\log(\lambda) \sim \chi^2(r)$$

with r equal to the difference in the number of constrained parameters between H_0 and H_1.

However, algebra can often be done to find a “nice” statistic.

This method often leads to usual χ^2, F, and t statistics.

LRT Example - Mean

Example: $y_i = \mu + \varepsilon_i$, where $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$, $i = 1,\ldots,n$

$\theta=(\mu, \sigma^2)'$
$g(\theta)=C\theta$
$C=(1,0)$
$\theta_0=\mu_0$

H_0: $\mu=\mu_0$, $\sigma^2>0$ vs H_1: $\mu\neq\mu_0$, $\sigma^2>0$

1) Writing the likelihood function of the observations.

$$L(\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

The log Likelihood is:

$$LL(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

LRT Example - Linear Regression

Example: $y = X\beta + \varepsilon$, where $\varepsilon \sim N(0,\sigma^2 I_n)$.

H_0: $C\beta=\gamma$, $\sigma^2>0$ vs H_1: $C\beta\neq\gamma$, $\sigma^2>0$

i.e. $\beta=(\beta_0,\beta_1)'$, $C=(0,1)$, $\gamma=0$ for

H_0: $\beta_1=0$, $\sigma^2>0$ vs H_1: $\beta_1\neq 0$, $\sigma^2>0$

$\beta=(\beta_0,\dots,\beta_q)'$
$g(\beta)=C\beta$
$C=(0,\dots,1)$
$\gamma=0$

1) Writing the likelihood function of the observations.

$$L(\beta,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

The log Likelihood is:

$$LL(\beta,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)$$

LRT Example - Difference in Means

Example:

$$y_{1i} = \mu_1 + \varepsilon_{1i} \qquad \varepsilon_{1i} \overset{iid}{\sim} N(0,\sigma_1^2) \qquad i=1,\ldots,n_1$$

$$y_{2j} = \mu_2 + \varepsilon_{2j} \qquad \varepsilon_{2j} \overset{iid}{\sim} N(0,\sigma_2^2) \qquad j=1,\ldots,n_2$$

H_0: $\mu_1 = \mu_2$ vs H_1: $\mu_1 \neq \mu_2$

1) Writing the likelihood function of the observations.

$$L(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2) = (2\pi\sigma_1^2)^{-\frac{n_1}{2}} e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2} (2\pi\sigma_2^2)^{-\frac{n_2}{2}} e^{-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{2j}-\mu_2)^2}$$

The log Likelihood is:

$$LL(\mu,\sigma^2) = -\frac{n_1+n_2}{2}\log(2\pi) - \frac{n_1}{2}\log(\sigma_1^2) - \frac{n_1}{2}\log(\sigma_2^2)$$

$$-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{1j}-\mu_2)^2$$

LRT Example - ANOVA

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$$

Example:

$$y_{1i} = \mu_1 + \varepsilon_{1i} \qquad \varepsilon_{1i} \overset{iid}{\sim} N(0,\sigma^2) \qquad i = 1,\ldots,n_1$$

$$y_{2j} = \mu_2 + \varepsilon_{2j} \qquad \varepsilon_{2j} \overset{iid}{\sim} N(0,\sigma^2) \qquad j = 1,\ldots,n_2$$

$$y_{3k} = \mu_3 + \varepsilon_{3k} \qquad \varepsilon_{3k} \overset{iid}{\sim} N(0,\sigma^2) \qquad k = 1,\ldots,n_3$$

r =3 populations

H_0: $\mu_1=\mu_2=\mu_3=\mu$ vs H_1: means not all same

1) Writing the likelihood function of the observations.

$$L(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2 \mid H_1) = (2\pi\sigma^2)^{-\frac{n_1}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2}$$

$$\times (2\pi\sigma^2)^{-\frac{n_2}{2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n_2}(y_{2j}-\mu_2)^2} (2\pi\sigma^2)^{-\frac{n_3}{2}} e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n_3}(y_{3k}-\mu_3)^2}$$

Multivariate Statistics

Daniel B. Rowe, Ph.D.

Associate Professor
Department of Mathematics,
Statistics, and Computer Science

MARQUETTE UNIVERSITY
Be The Difference.

Outline

- Bi(Multi)variate Normal Distribution
- Wishart Distribution
- Multivariate Student t
- Matrix Normal and Matrix T

Bivariate Normal Distribution

If a random variable $\underset{2\times 1}{x}$ has a normal distribution with

mean vector $\underset{2\times 1}{\mu}$ and variance-covariance matrix Σ, then

$$f(x|\mu,\Sigma)=(2\pi)^{-p/2}|\Sigma|^{-1/2}e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

(annotations: Σ – covariance matrix; μ – mean vector; Σ^{-1} – covariance matrix)

$x,\mu\in\mathbb{R}^p$

$p=2$

$\Sigma>0$ ← set of pos def matrices

and we write $\underset{2\times 1}{x}\sim N(\mu,\Sigma)$ (annotations: μ – mean vector; Σ – covariance matrix). The covariance matrix $\underset{2\times 2}{\Sigma}$, has to

be of full rank.

Bivariate Normal Distribution

$$z = B(x-\mu)$$

The distribution of vector variable $\underset{2\times 1}{x}$ (joint of x_1 and x_2) is

$$f_X(x|\theta) = f_Z(z(x)) \times |J(z \to x)|$$

$$f(z) = (2\pi)^{-2/2} |I_n|^{-1/2} e^{-\frac{1}{2}(z-0)'(I_n)^{-1}(z-0)} \qquad J = \frac{\partial z}{\partial x} = B$$

$$f_X(x|\mu,\Sigma) = (2\pi)^{-2/2} |I_n|^{-1/2} e^{-\frac{1}{2}(B(x-\mu)-0)'(I_n)^{-1}(B(x-\mu)-0)} |B|$$

$$f_X(x|\mu,\Sigma) = (2\pi)^{-2/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \qquad x,\mu \in \mathbb{R}^p$$

$$\Sigma > 0$$

Bivariate Normal Distribution

Theorem:

If x is a 2-D (or p-D) random variable from $f(x|\mu,\Sigma)$, with

$$E(x|\mu,\Sigma)=\mu$$

$$\text{var}(x|\mu,\Sigma)=\Sigma$$

then we form $y=Ax+\delta$ where dimensions match

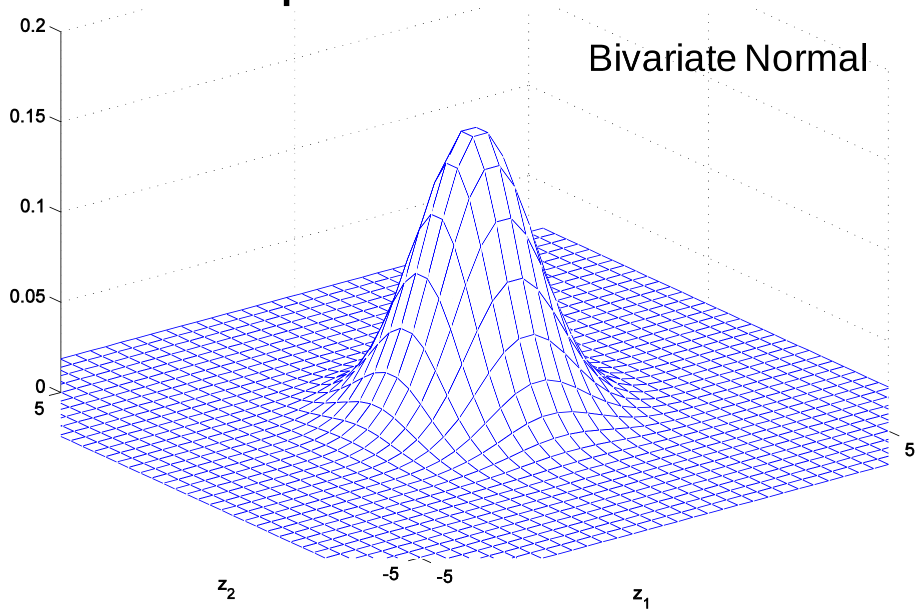
and A full row rank (A: $r\times p$, $r\leq p$), then

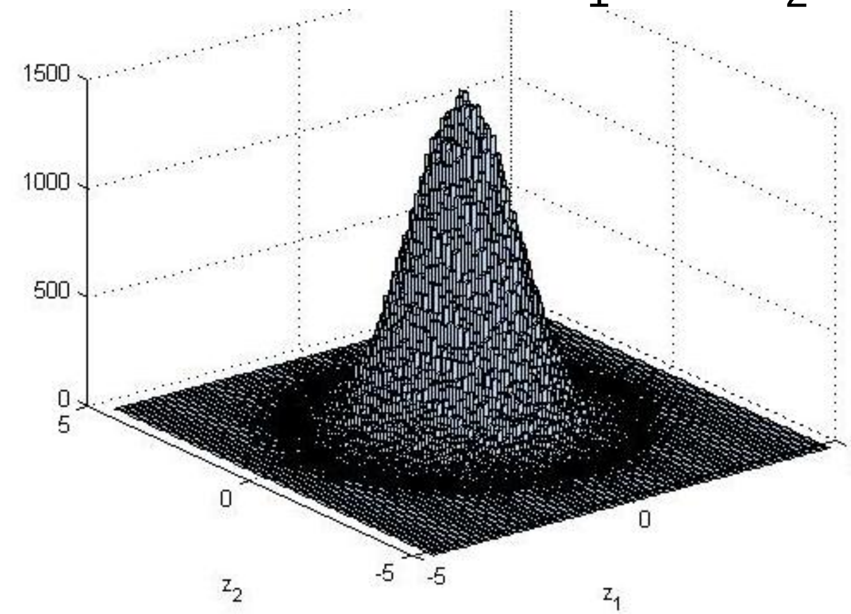
$$E(y|\mu,\Sigma,\delta,A)=A\mu+\delta$$

$$\text{var}(y|\mu,\Sigma,A)=A\Sigma A'.$$

Bivariate Normal Distribution

Obtain 2-D standard normal variates $\underset{2\times 1}{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ by transforming two independent standard uniform random variates u_1 and u_2.





First half of 10^6 standard normal variates as z_1's and second half as z_2's. Produce 5×10^5 statistically independent z's.

Bivariate Normal Distribution

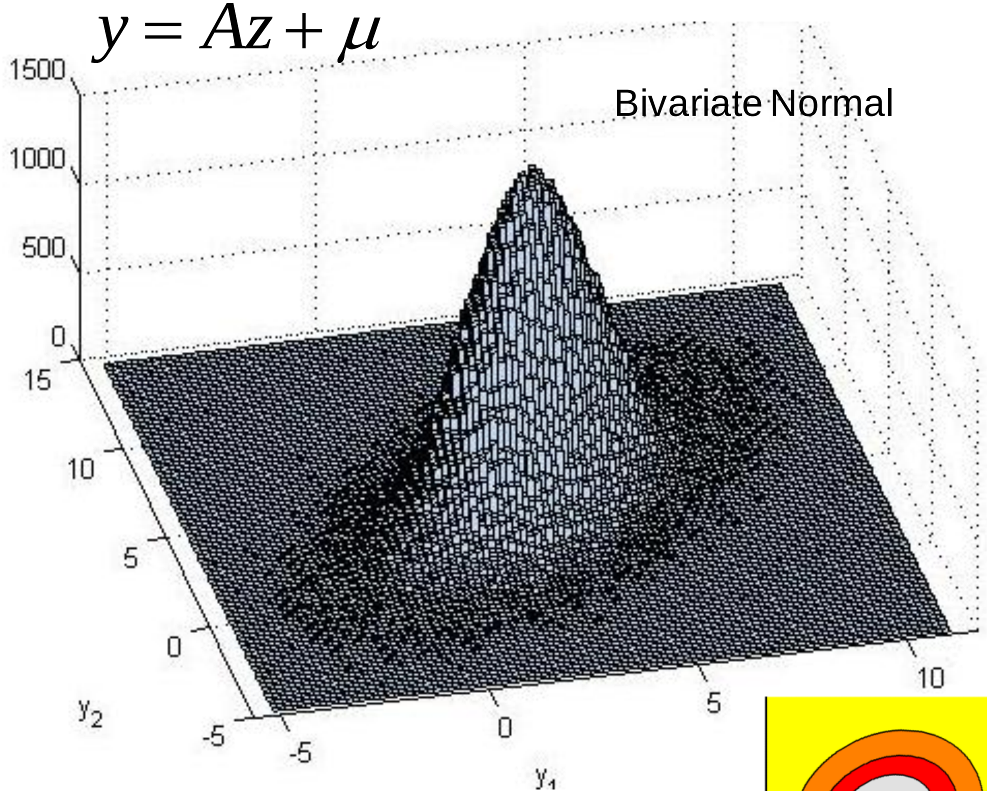
$$\mu = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$$

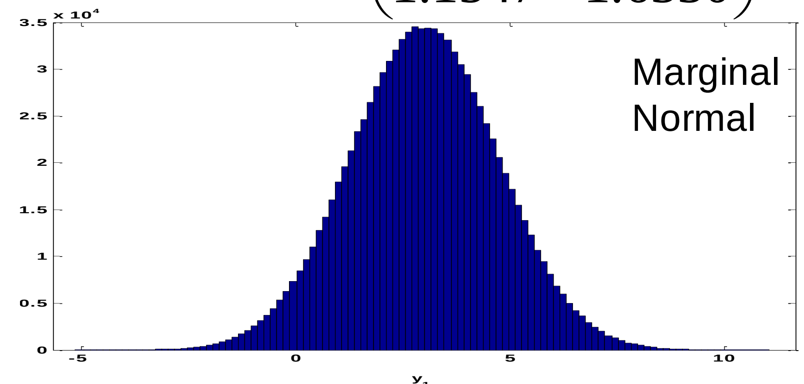
Multiplied 5×10^5 simulated z's by A and added μ. $\rho = 0.58$

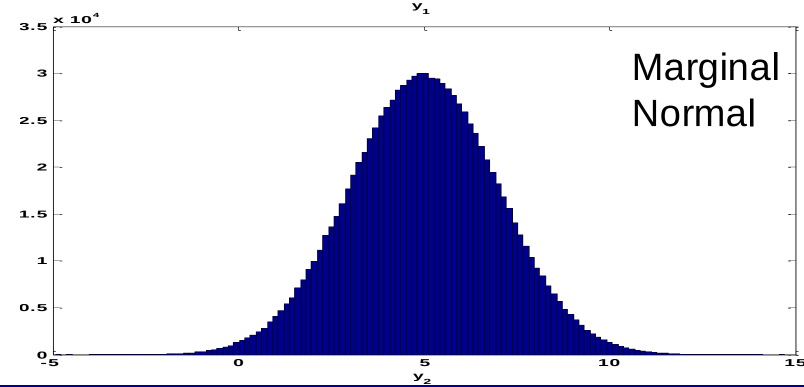
The y's are now $N(\mu, \Sigma = AA')$.

$$y = Az + \mu$$

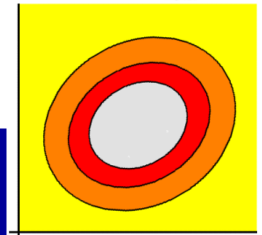
Cholesky $$A = \begin{pmatrix} 1.7321 & 0 \\ 1.1547 & 1.6330 \end{pmatrix}$$







cross section



Bivariate Normal Distribution

Theorem:

If x_1 and x_2 are independent 2-D (or any D) RVs with

$E(x_1 | \mu_1, \Sigma_1) = \mu_1$ $\quad$ $E(x_2 | \mu_2, \Sigma_2) = \mu_2$

$\text{var}(x_1 | \mu_1, \Sigma_1) = \Sigma_1$ $\quad$ $\text{var}(x_2 | \mu_2, \Sigma_2) = \Sigma_2$

then if we let $y = (x_1 + x_2)/2$,

$$E(y | \mu_1, \Sigma_1, \mu_2, \Sigma_2) = (\mu_1 + \mu_2)/2$$

$$\text{var}(y | \mu_1, \Sigma_1, \mu_2, \Sigma_2) = (\Sigma_1 + \Sigma_2)/4$$

Bivariate Normal Distribution

Theorem:

If $x_1,\ldots,x_n$ are independent 2-D (or any D) RVs with

$$E(x_i \mid \mu,\Sigma)=\mu \qquad i=1,\ldots,n$$

$$\text{var}(x_i \mid \mu,\Sigma)=\Sigma$$

then if we let $\bar{x}=(x_1+\cdots+x_n)/n$,

$$E(\bar{x} \mid \mu,\Sigma)=\mu$$

$$\text{var}(\bar{x} \mid \mu,\Sigma)=\Sigma/n$$

If x's are normal, then $\bar{x} \sim N(\mu,\Sigma/n)$.

Bivariate Normal Distribution

If $x_1,\ldots,x_n$ are IID $N(\mu,\Sigma)$ 2-D (or any D) RVs and

$$\bar{x} = (x_1 + \cdots + x_n)/n \sim N(\mu, \Sigma/n)$$

then

$$f_{\bar{X}}(\bar{x} \mid \mu, \Sigma) = (2\pi)^{-2/2} \, |\Sigma/n|^{-1/2} \, e^{-\frac{1}{2}(\bar{x}-\mu)'(\Sigma/n)^{-1}(\bar{x}-\mu)} .$$

Also note that

$$n^{1/2}\Sigma^{-1/2}(\bar{x}-\mu) \sim N(0, I_2) . \qquad \Sigma^{-1/2} = A^{-1} = B$$

Estimate Σ by $\hat{\Sigma} = n^{1/2} \sum_{i=1}^{n} (x_i - \bar{x})'(x_i - \bar{x})$.

Wishart Distribution

Multivariate version of gamma distribution.

A random $p \times p$ matrix variate $\underset{p\times p}{G}$ follows the Wishart distribution with scale matrix $\underset{p\times p}{\Sigma}$ and ν df denoted $\underset{p\times p}{G} \sim W(\underset{p\times p}{\Sigma}, \nu)$

$$G, \Sigma > 0$$
$$\nu > p+1$$
$$\nu \in \mathbb{N}$$

iff $f(G \mid \Sigma, \nu) = k_W \, |\Sigma|^{-\frac{\nu}{2}} |G|^{\frac{\nu-p-1}{2}} e^{-\frac{1}{2} tr(\Sigma^{-1} G)}$

where $k_W^{-1} = 2^{\frac{\nu p}{2}} \pi^{-\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\frac{\nu+1-j}{2}\right)$

If p=1, $f(g \mid \sigma^2, \nu) = \dfrac{g^{\frac{\nu}{2}-1} e^{-\frac{g}{2\sigma^2}}}{\Gamma(\nu/2)(2\sigma^2)^{\nu/2}}$

Gamma dist by $\alpha = \nu / 2$, Chi-square by $y = g / \sigma^2$.

Wishart Distribution

$$f(G|\Sigma,\nu) = k_W |\Sigma|^{-\frac{\nu}{2}} |G|^{\frac{\nu-p-1}{2}} e^{-\frac{1}{2}tr(\Sigma^{-1}G)}$$

$$k_W^{-1} = 2^{\frac{\nu p}{2}} \pi^{-\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\frac{\nu+1-j}{2}\right)$$

The mean, variance, and covariance of elements are

$$E(G|\Sigma,\nu) = \nu\Sigma$$

$$\text{var}(G_{ij}|\Sigma,\nu) = \nu(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj})$$

$$\text{cov}(G_{ij}G_{kl}|\Sigma,\nu) = \nu(\Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk})$$

Wishart Distribution

We obtain a (singular) $W(\Sigma,1)$ variate G_1 by transforming a centered normal vector variate $z_1=(x_1-\mu)$ via $G_1=z_1z_1'$ and a $W(\Sigma,n)$ variate by $G=\sum_{i=1}^{n} z_iz_i'$. If the mean vector is not known, then we can estimate it and lose one *df*.

$$\underbrace{\sum_{i=1}^{n}(x_i-\mu)'\overset{G}{(x_i-\mu)}}_{W(\Sigma,n),n\geq p}=\underbrace{\sum_{i=1}^{n}(x_i-\bar{x})'\overset{G_2}{(x_i-\bar{x})}}_{W(\Sigma,n-1),n-1>p}+\underbrace{n(\bar{x}-\mu)'\overset{G_1}{(\bar{x}-\mu)}}_{W(\Sigma,1),\text{ singular}}$$

add and subtract $\bar{x}$ in parentheses

Let p=1 and get σ^2 times χ^2 result.

Wishart Distribution

Theorem:

If G is a p×p random matrix variable from $f(G|\Sigma,\nu)$, with

$$f(G|\Sigma,\nu)=k_W\,|\Sigma|^{-\frac{\nu}{2}}|G|^{\frac{\nu-p-1}{2}}\,e^{-\frac{1}{2}tr(\Sigma^{-1}G)}$$

then if we form $Q=AGA'$ where dimensions match

and A full row rank (A: r×p, r≤p), then $Q\sim W(\Delta=A\Sigma A',\nu)$

$$E(G|\Delta,\nu)=\nu\Delta$$

$$\text{var}(Q_{ij}|\Delta,\nu)=\nu(\Delta_{ij}^2+\Delta_{ii}\Delta_{jj})$$

$$\text{cov}(G_{ij}G_{kl}|\Delta,\nu)=\nu(\Delta_{ik}\Delta_{jl}+\Delta_{il}\Delta_{jk}).$$

Wishart Distribution

Took 5×10^4 sets of n=10 variates x, subtracted mean $\mu=\begin{pmatrix}3\\5\end{pmatrix}$ from each set, transpose multiplied each value, added the 10 values in set to form each G. The G's are now $W(\Sigma,\nu=n)$.

$$E(G\mid\Sigma,\nu)=\nu\Sigma \qquad \nu=10 \quad \Sigma=\begin{pmatrix}3&2\\2&4\end{pmatrix}$$

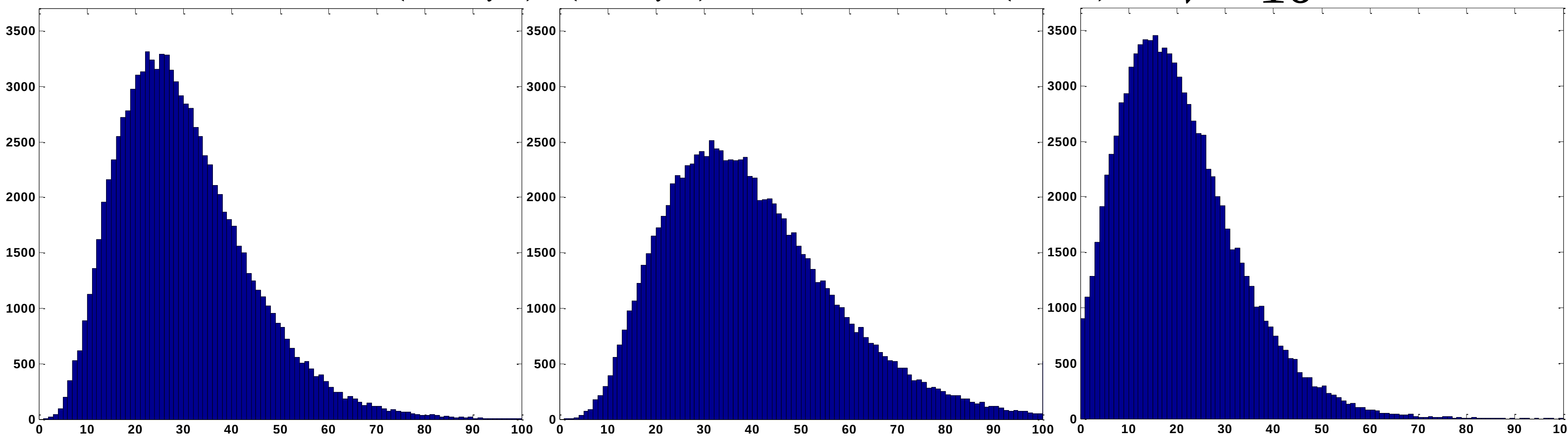
$$\operatorname{var}(G_{ij}\mid\Sigma,\nu)=\nu(\Sigma_{ij}^2+\Sigma_{ii}\Sigma_{jj})$$

$$\operatorname{cov}(G_{ij}G_{kl}\mid\Sigma,\nu)=\nu(\Sigma_{ik}\Sigma_{jl}+\Sigma_{il}\Sigma_{jk})$$

Wishart Distribution

$$\mu = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$$

The G's, $G = (x-\mu)'(x-\mu)$ are now $W(\Sigma,\nu)$. $\nu = 10$



$G(1,1)$'s $G(2,2)$'s $G(1,2)$'s

$$E(G \mid \Sigma,\nu) = \nu\Sigma = \begin{pmatrix} 30 & 20 \\ 20 & 40 \end{pmatrix} \qquad \mathrm{var}(G_{ij} \mid \Sigma,\nu) = \nu(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}) = \begin{pmatrix} 180 & 160 \\ 160 & 320 \end{pmatrix}$$

$$\mathrm{cov}(G_{ij}G_{kl} \mid \Sigma,\nu) = \nu(\Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}) = 80, 120, 160$$

$11,22 \quad 11,12 \quad 22,12 \leftarrow ij,kl$

Wishart Distribution

Took 5×10^4 sets of n=10 variates x, subtracted mean $\bar{x}$ from each set, transpose multiplied each value, added the 10 values to form each G_2. The G_2's are now $W(\Sigma,\nu=n-1)$.

$$E(G|\Sigma,\nu)=\nu\Sigma \qquad \nu=9 \quad \Sigma=\begin{pmatrix}3 & 2\\ 2 & 4\end{pmatrix}$$

$$\mathrm{var}(G_{ij}|\Sigma,\nu)=\nu(\Sigma_{ij}^2+\Sigma_{ii}\Sigma_{jj})$$

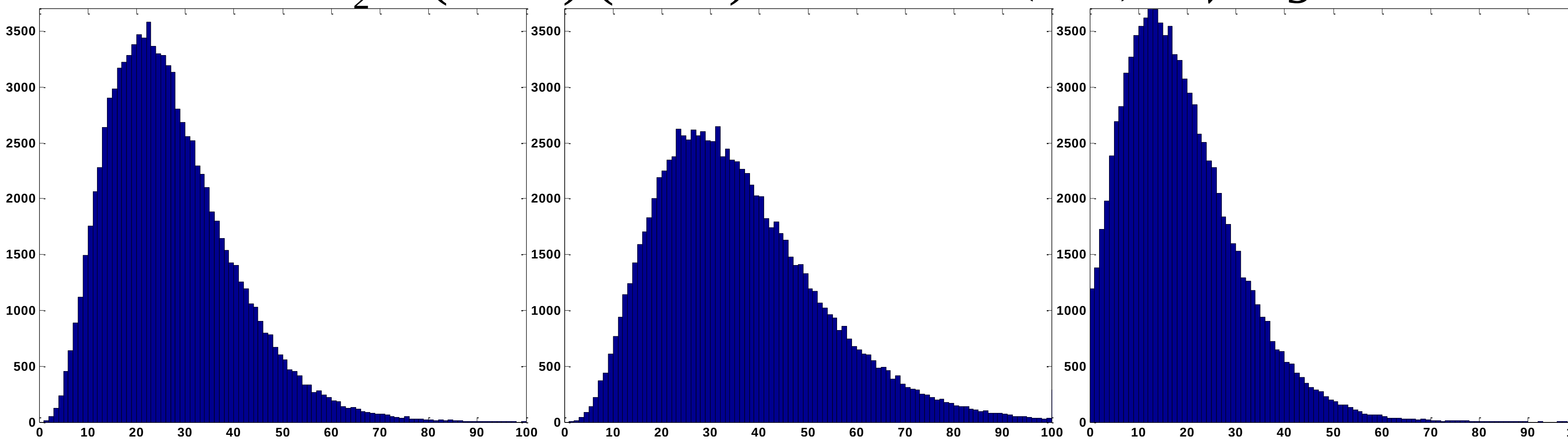
$$\mathrm{cov}(G_{ij}G_{kl}|\Sigma,\nu)=\nu(\Sigma_{ik}\Sigma_{jl}+\Sigma_{il}\Sigma_{jk})$$

Wishart Distribution

$\mu = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$

Toggle with Next slide

The G_2's, $G_2 = (x-\bar{x})'(x-\bar{x})$ are now $W(\Sigma,\nu)$. $\nu = 9$



$G_2\,(1,1)$'s $\qquad G_2\,(2,2)$'s $\qquad G_2\,(1,2)$'s

$$E(G_2 \mid \Sigma,\nu) = \nu\Sigma = \begin{pmatrix} 27 & 18 \\ 18 & 36 \end{pmatrix} \qquad \mathrm{var}(G_{2ij} \mid \Sigma,\nu) = \nu\,(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}) = \begin{pmatrix} 162 & 144 \\ 144 & 288 \end{pmatrix}$$

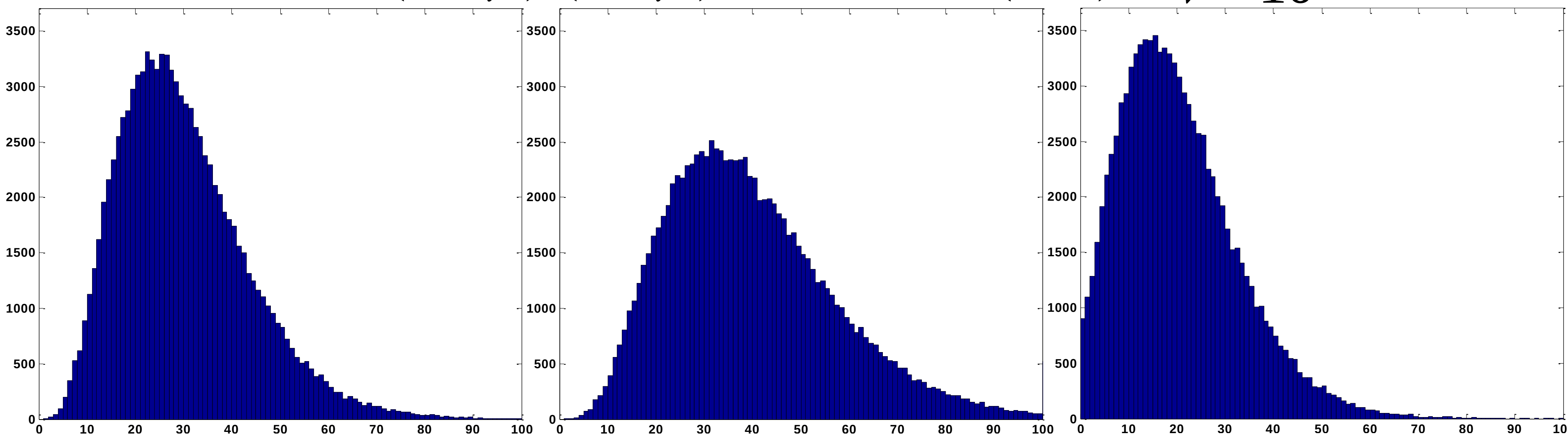
$$\mathrm{cov}(G_{2ij}G_{2kl} \mid \Sigma,\nu) = \nu(\Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}) = \underset{11,22}{72}\,,\underset{11,12}{108},\underset{22,12}{144} \leftarrow ij,kl$$

Wishart Distribution

$$\mu=\begin{pmatrix}3\\5\end{pmatrix}\quad \Sigma=\begin{pmatrix}3 & 2\\ 2 & 4\end{pmatrix}$$

Toggle with
Previous slide

The G's, $G=(x-\mu)'(x-\mu)$ are now $W(\Sigma,\nu)$. $\nu=10$



$G(1,1)$'s $G(2,2)$'s $G(1,2)$'s

$$E(G\,|\,\Sigma,\nu)=\nu\Sigma=\begin{pmatrix}30 & 20\\ 20 & 40\end{pmatrix}\qquad \mathrm{var}(G_{ij}\,|\Sigma,\nu)=\nu(\Sigma_{ij}^2+\Sigma_{ii}\Sigma_{jj})=\begin{pmatrix}180 & 160\\ 160 & 320\end{pmatrix}$$

$$\mathrm{cov}(G_{ij}G_{kl}\,|\Sigma,\nu)=\nu(\Sigma_{ik}\Sigma_{jl}+\Sigma_{il}\Sigma_{jk})=\underset{11,22}{80},\underset{11,12}{120},\underset{22,12}{160}\leftarrow ij,kl$$

Multivariate Student-t Distribution

A 2-D (p=2) random vector variate $\underset{p\times 1}{t}$ follows a standard

Student t distribution with ν df denoted $\underset{p\times 1}{t} \sim t(\nu)$ iff

$$f(t\,|\,\nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{(\nu\pi)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left[1+\frac{1}{\nu}t't\right]^{\frac{\nu+p}{2}}}$$, where $t \in \mathbb{R}^p$, $\nu \in \mathbb{N}$.

The mean and variance of t are

$$E(t\,|\,\nu) = \underset{p\times 1}{0}$$

$$\nu > 1$$

$$\text{var}(t\,|\,\nu) = \frac{\nu}{\nu-2} \underset{p\times p}{I_p}$$

$$\nu > 2$$

Multivariate Student-t Distribution

We obtain a standard $t(v=n$-1) variate t by transforming a

normal 2-D (p-D) random variate $\bar{x}$ and a Wishart

random matrix variate G that are independent

$$\begin{pmatrix} t \\ V \end{pmatrix} = \begin{pmatrix} n^{1/2} v^{1/2} G_2^{-1/2} (\bar{x} - \mu) \\ G_2 \end{pmatrix} \qquad \underset{p\times 1}{\bar{x}} \sim N(\mu, \Sigma / n) \quad \underset{p\times p}{G_2} \sim W(\Sigma, v)$$

.

The original variables in terms of the new variables

$$\begin{pmatrix} \bar{x}(t,v) \\ G_2(t,V) \end{pmatrix} = \begin{pmatrix} n^{-1/2} v^{-1/2} V^{1/2} t + \mu) \\ V \end{pmatrix}$$

Multivariate Student-t Distribution

The joint distribution of (t,V) can be obtained as

$$f(t,V\,|\,\mu,\Sigma,\nu)=f(\bar{x}(t,V),G_2(t,V))\times|\,J(\bar{x},G_2\to t,V)\,|$$

The Jacobian of the transformation is

$$|\,J(\bar{x},G_2\to t,V)\,|=n^{-1/2}\nu^{-1/2}V^{1/2}$$

The joint PDF of (t,V) is

$$p(t,V\,|\,\Sigma,\nu)=(2\pi\nu)^{-p/2}k_W\,|\,\Sigma\,|^{-\frac{\nu+1}{2}}|\,V\,|^{-\frac{(\nu+1)-p-1}{2}}\;e^{-\frac{1}{2}tr\Sigma^{-1}V(I_p+\frac{1}{\nu}tt')}$$

The distribution of t can be found by integrating out V

$$f(t\,|\,\nu)=\frac{\Gamma\left(\frac{\nu+p}{2}\right)}{(\nu\pi)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\left[1+\frac{1}{\nu}t't\right]}\;.$$

After integration use

$$|\,I_p+\tfrac{1}{\nu}tt'\,|=|\,1+\tfrac{1}{\nu}t't\,|$$

Multivariate Student-t Distribution

Recall $\bar{x} \sim N(\mu,\Sigma/n)$

so $n^{1/2}\Sigma^{-1/2}(\bar{x}-\mu) \sim N(0,I)$

also $\sum_{i=1}^{n}(x_i-\bar{x})'(x_i-\bar{x}) \sim W(\Sigma,n-1)$

thus $(n-1)\Sigma^{-1/2}\left[\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i-\bar{x})'(x_i-\bar{x})\right](\Sigma^{-1/2})' \sim W(I,n-1)$

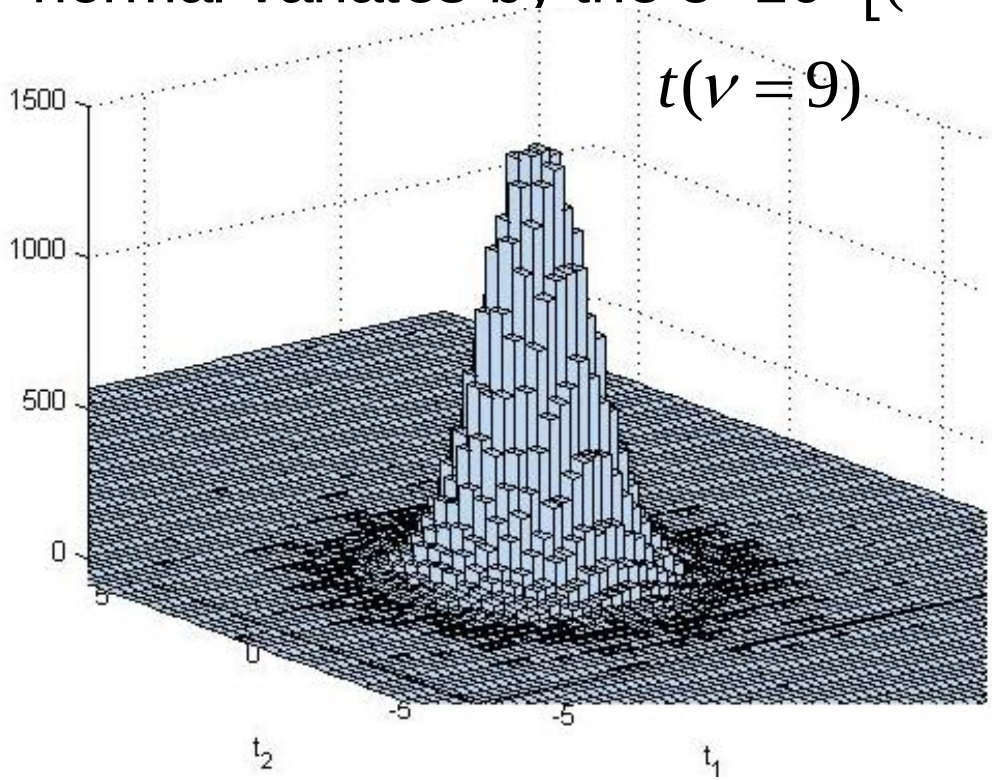
but $\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i-\bar{x})'(x_i-\bar{x}) = S$

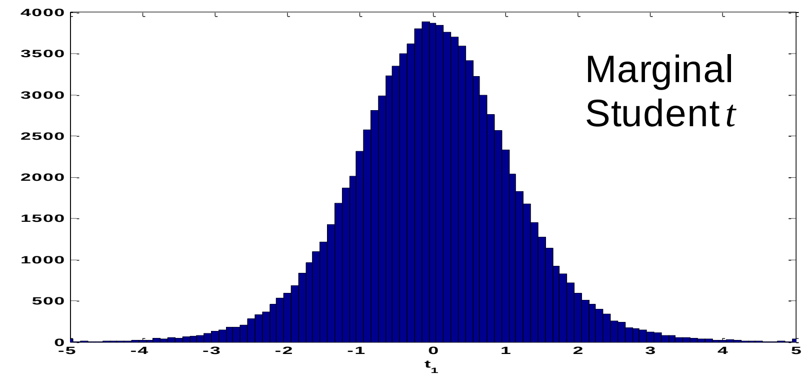
finally $[(n-1)S]^{-1/2} n^{1/2}(\bar{x}-\mu) \sim t(n-1)$

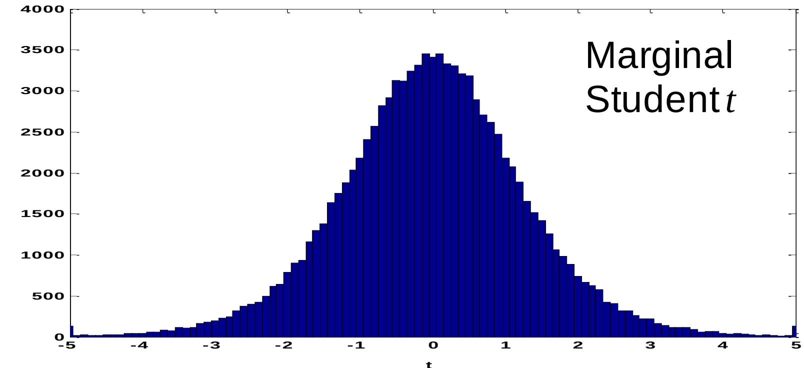
Multivariate Student-t Distribution

$$\mu = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$$

Take sample mean $\bar{x}$ from each 5×10⁴ sets of 10 variates, subtracted μ from each set, premultiplied the 5×10⁴ centered normal variates by the 5×10⁴ $\left[(n-1)S\right]^{-1/2} n^{1/2}(\bar{x}-\mu)$.







Multivariate Student-t Distribution

A 2-D (p=2) random vector variate $\underset{p\times 1}{s}$ follows a general

Student t distribution with location $\underset{p\times 1}{\mu}$ and scale $\underset{p\times p}{\Sigma}$ and df ν

$$s \sim t(\nu, \mu, \Sigma)$$

$$f(s \mid \nu, \mu, \Sigma) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{(\nu\pi)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left[1+\frac{1}{\nu}(s-\mu)'\Sigma^{-1}(s-\mu)\right]^{\frac{\nu+p}{2}}}$$

where $s, \mu \in \mathbb{R}^p$, $\nu \in \mathbb{N}$, $\Sigma > 0$.

The mean and variance of s are

$$E(t \mid \nu, \mu, \Sigma) = \underset{p\times 1}{\mu} \qquad \nu > 1 \qquad\qquad \text{var}(s \mid \nu, \mu, \Sigma) = \frac{\nu}{\nu-2}\underset{p\times p}{\Sigma} \qquad \nu > 2$$

Multivariate Distributions

We can also perform hypothesis tests.

Single mean $H_o : \underset{p\times 1}{\mu} = \underset{p\times 1}{\mu_0}$ vs $H_1 : \underset{p\times 1}{\mu} \neq \underset{p\times 1}{\mu_0}$

Two Means $H_o : \underset{p\times 1}{\mu_1} = \underset{p\times 1}{\mu_2}$ vs $H_1 : \underset{p\times 1}{\mu_1} \neq \underset{p\times 1}{\mu_2}$

MANOVA $H_o : \underset{p\times 1}{\mu_1} = \underset{p\times 1}{\mu_2} = \underset{p\times 1}{\mu_3}$ vs $H_1 : \mu$'s not all equal

Multivariate Regression $\underset{n\times p}{Y} = \underset{n\times (q+1)}{X} \overset{(q+1)\times p}{B} + \underset{n\times p}{E}$ $\hat{B} = (X'X)^{-1}X'Y$

$S = \frac{1}{n-q-1}(Y - X\hat{B})'(Y - X\hat{B})$

$H_o : \underset{p\times p}{\Sigma} = \underset{p\times p}{diagonal}$ vs $H_1 : \underset{p\times p}{\Sigma} \neq \underset{p\times p}{diagonal}$

Take home Exam posted tomorrow!