Sampling from the normal

Normal samples

1. Sometimes data is normal.
2. Large sample statistics are often normal even when the underlying data isn't (central limit theorem).
3. The results are pretty cute.
4. It's a *tradition!*

Properties of sample mean and variance

Theorem 5.3.1:

Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution and let $\bar{x} = {}^1/_n \sum_{i=1}^{n} x_i$ and $s^2 = {}^1/_{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$. Then,

a) $\bar{x}$ and s^2 are independent random variables

b) $\bar{x} \sim N(\mu, \sigma^2/n\,)$ (We already showed this, since the sample mean is a linear combination of normal random variables.)

c) $(n-1)s^2/\sigma^2$ has a chi-squared distribution with $n-1$ degrees of freedom (χ^2_{n-1} or $\chi^2(n-1)$).

The source of all proofs about linear models is *Linear Models, S.R.* Searle, Wiley, 1971.

Quadratic and linear forms

Theorem:

If $x \sim N(\mu, \Sigma)$, then $x'Ax$ and Bx, where A is symmetric positive definite, are distributed independently if and only if $B\Sigma A = 0$.

Outline of proof

- Prove sufficiency by cleverly messing with matrix square roots.
- Digression on algebra of quadratic form
- Prove necessity
  - Lemma on traces
  - Lemma on covariances

Sufficiency $B\Sigma A = 0$

We can write $A = LL'$ for some L of full column rank, which implies $(L'L)^{-1}$ exists. Note $B\Sigma A = B\Sigma LL'$

$$B\Sigma A = 0 \Rightarrow B\Sigma LL' = 0$$

$$B\Sigma LL'[L(L'L)^{-1}] = 0[L(L'L)^{-1}]$$

$$B\Sigma L = 0$$

Consider Bx and $L'x$, these are joint normal with covariance $B\Sigma L = 0$ and therefore independent. Since

$$x'Ax = (L'x)'(L'x) = g(L'x)$$

This proves $x'Ax$ and Bx are independent.

Lemmas needed for necessity

Lemma (stated without proof):

$$\mathrm{E}(x'Ax) = \mathrm{tr}(A\Sigma) + \mu' A\mu$$

where tr( ) is the trace, that is the sum of the diagonal elements.

Digression

$$(x-\mu)'A(x-\mu) = x'Ax + \mu'A\mu - 2x'A\mu$$

$$x'Ax - \mu'A\mu$$
$$= (x-\mu)'A(x-\mu) - 2\mu'A\mu + 2x'A\mu$$

Lemmas needed for necessity

Lemma:

$$\mathrm{cov}(x, x'Ax) = 2\Sigma A\mu$$

Proof:

$$\mathrm{cov}(x, x'Ax) = \mathrm{E}\big((x-\mu)[x'Ax - \mathrm{E}(x'Ax)]\big)$$
$$= \mathrm{E}\big((x-\mu)[x'Ax - \mu'A\mu - \mathrm{tr}(A\Sigma)]\big)$$

Insert digression results

$$= \mathrm{E}\big((x-\mu)[(x-\mu)'A(x-\mu) + 2(x-\mu)'A\mu - \mathrm{tr}(A\Sigma)]\big)$$

Multiply through by the first $(x-\mu)$. First term is essentially third moment. Last term is proportional to first moment. Both are zero.

$$\mathrm{cov}(x, x'Ax) = 2\Sigma A\mu$$

Necessity $x'Ax$ and Bx independent

$x'Ax$ and Bx independent implies covariance equals zero, so

$$\mathrm{cov}(Bx, x'Ax) = 2B\Sigma A\mu = 0$$

Since this is true for all μ it must be that $B\Sigma A = 0$.

Special case

Let $\Sigma = \sigma^2 I_n$, so $x \sim N(\mu, \Sigma) = N(\mu, \sigma^2 I)$.

Let $B = \frac{1}{n}[1 \ldots 1]$, then $\bar{x} = Bx$.

If we let

$$A = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix}$$

$$s^2 = \frac{1}{n-1} x'Ax$$

$n = 2$ case

$$x'Ax = [x_1 \quad x_2]\begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \left[x_1 - \frac{x_1 + x_2}{2} \quad x_2 - \frac{x_1 + x_2}{2}\right]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= x_1^2 + x_2^2 - (x_1 + x_2)\frac{x_1 + x_2}{2}$$

$$= \left(x_1 - \frac{x_1 + x_2}{2}\right)^2 + \left(x_2 - \frac{x_1 + x_2}{2}\right)^2$$

Special case

$$B\Sigma A = B\sigma^2 IA$$

But $BA = 0$

$$\frac{1}{n}[1 \dots 1]\begin{bmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{bmatrix} = 0$$

Wow!

Therefore, the sample mean and the sample variance of normals are independent.

Reminders about χ^2

Lemma 5.3.2:

a) If $z \sim N(0,1)$, then $z^2 \sim \chi_1^2$.
b) If $z \sim N(0, \sigma^2)$, then $\frac{z}{\sigma} \sim N(0,1)$, so $\frac{1}{\sigma^2} z^2 \sim \chi_1^2$.
c) If $x_1, \dots, x_n$ are independent and $x_i \sim \chi_{p_i}^2$, then $x_1 + \cdots + x_n \sim \chi_{p_1 + \cdots + p_n}^2$.

Recap- Properties of sample mean and variance

Theorem 5.3.1:
Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution and let $\bar{x} = {}^1/_n \sum_{i=1}^{n} x_i$ and $s^2 = {}^1/_{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$. Then,

a) $\bar{x}$ and s^2 are independent random variables
b) $\bar{x} \sim N(\mu, \sigma^2/n)$ (We already showed this, since the sample mean is a linear combination of normal random variables.)
c) $(n-1)s^2/\sigma^2$ has a chi-squared distribution with $n-1$ degrees of freedom (χ^2_{n-1} or $\chi^2(n-1)$).

Properties of s^2

If $x \sim \chi^2_k$

$$\mathrm{E}(x) = k$$

$$\mathrm{E}\left(\frac{\sigma^2}{n-1} \times \frac{(n-1)s^2}{\sigma^2}\right) = \frac{\sigma^2}{n-1} \times (n-1) = \sigma^2$$

$$\mathrm{var}(x) = 2k$$

$$\mathrm{var}\left(\frac{\sigma^2}{n-1} \times \frac{(n-1)s^2}{\sigma^2}\right) = \left(\frac{\sigma^2}{n-1}\right)^2 \times 2(n-1)$$

$$= \frac{2\sigma^4}{n-1}$$

Assignment

Generate $n = 4$ standard normals, compute the sample mean $\bar{x}$, sample variance s^2, and the t-statistic given by $t = \bar{x}/\sqrt{s^2/n}$. Do this a lot of times, saving the result. (While you're at it, use tic/toc to compute how long your simulation takes.) Now make three plots showing the empirical and theoretical pdfs, one for the sample mean, one for $(n-1)$ times the sample variance, and one for the t-statistic.

Useful matrix theorems

Definition:
A is positive definite iff the quadratic form $c'Ac > 0 \ \forall c \neq 0$.

1. If A is symmetric, positive definite, $\exists$ nonsingular P s.t. $A = PP'$
2. If A is $n \times k$, $n \geq k$, rank k, then $A'A$ is symmetric pos. def. If $rank(A) < k$, $A'A$ is positive semi-definite.
3. If A is of full column rank, then $\det(A'A) > 0$.
4. If A, B are pos. def. and $A - B$ is pos. def., then $B^{-1} - A^{-1}$ is pos. def.

Linear and quadratic forms

If $u \sim (\mu, \Sigma)$, then $Lu \sim (L\mu, L\Sigma L')$, where u is $n \times 1$, L is $q \times n$.

1. If $u \sim N(0, \sigma^2 I)$ and A is symmetric, idempotent, rank r, then $\frac{1}{\sigma^2} u'Au \sim \chi_r^2$.
2. If $u \sim N(0, V)$ then $u'Au \sim \chi_r^2$, iff AV is idempotent of rank r.
3. (Corollary) If $u \sim N(0, V)$ then $u'V^{-1}u \sim \chi_r^2$, where $rank(V) = r$.
4. If A, B sym. idem. and $u \sim N(0, \sigma^2 I)$, then $\frac{1}{\sigma^2} u'Au, \frac{1}{\sigma^2} u'Bu$ are independent iff $AB = 0$.
5. If $u \sim N(0, \sigma^2 I)$, $u'Au$, Bu are independent iff $BA = 0$.

Getting to *t*-

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

We *standardize* $\bar{x}$.

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

$$t = \frac{z}{\sqrt{s^2/\sigma^2}} = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} \sim t_{n-1}$$

Remember that s^2/σ^2 is a χ_{n-1}^2 divided by $n-1$. The ratio of a standard normal to the square root of an independent χ^2 divided by its degrees of freedom is distributed *t*-.

F-statistic

If $x_1 \sim \chi_{d_1}^2, x_2 \sim \chi_{d_2}^2$ and x_1, x_2 are independent, then

$$\frac{x_1/d_1}{x_2/d_2} \sim F(d_1, d_2)$$

Example:

$$t^2 = \frac{z^2/1}{s^2/\sigma^2} \sim F(1, n-1)$$

Another way

$$SSR = (n-1)s^2 = \sum (x - \bar{x})^2 = x'Ax, \qquad A = \begin{bmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{bmatrix}$$

Suppose we believe that $\mu = 0$, then

$$SSR^* = \sum x^2 = x'I_n x$$

Divided by σ^2 is also χ^2.

$$SSR^* - SSR = \frac{1}{n} x'Bx, B = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

So $SSR^* - SSR$ is also χ^2. Note that $BA = 0$, proving $SSR^* - SSR$ and SSR are independent. Therefore

$$\frac{(SSR^* - SSR)/1}{SSR/(n-1)} = \frac{(\sum x^2 - \sum (x - \bar{x})^2)/1}{\sum (x - \bar{x})^2/(n-1)} \sim F(1, n-1)$$

Least squares

$$y = X\beta + \varepsilon$$
$$\varepsilon \sim N\left(0, \sigma^2 I\right)$$
$$y \sim N\left(X\beta, \sigma^2 I\right)$$
$$\hat{\beta} = \left(X'X\right)^{-1} X'y$$
$$\hat{\beta} \sim N\left(\left[\left(X'X\right)^{-1} X'\right] X\beta, \left[\left(X'X\right)^{-1} X'\right] \sigma^2 I \left[\left(X'X\right)^{-1} X'\right]'\right)$$
$$\hat{\beta} \sim N\left(\beta, \sigma^2 \left(X'X\right)^{-1}\right)$$

Residuals

$$e = y - X\hat{\beta}$$
$$e = X\beta + \varepsilon - X\left(X'X\right)^{-1} X'(X\beta + \varepsilon)$$
$$e = \varepsilon - X\left(X'X\right)^{-1} X'\varepsilon$$
$$e = \left[I - X\left(X'X\right)^{-1} X'\right] \varepsilon$$
$$SSR = e'e = \varepsilon' \left[I - X\left(X'X\right)^{-1} X'\right] \varepsilon$$
$$\frac{SSR}{\sigma^2} \sim \chi^2_{n-k}$$
$$\left(X'X\right)^{-1} X'[I - X(X'X)^{-1} X'] = 0$$

$\hat{\beta}, s^2$ are independent