Random Vectors

x is a p×1 random vector with a pdf probability density function f(**x**): $\mathbf{R}^p \to \mathbf{R}$. Many books write **X** for the random vector and **X**=**x** for the realization of its value.

E[**X**]= $\int \mathbf{x}\, f(\mathbf{x})\, d\mathbf{x} = \mu$.

Theorem: E[A**x**+**b**]= AE[**x**]+**b**

Covariance Matrix E[(**x**-μ)(**x**-μ)']=var(**x**)=Σ (note the location of transpose)

Theorem: Σ=E[**xx**']-μμ'

If **y** is a random variable: covariance C(**x**,**y**)= E[(**x**-μ)(**y**-ν)']

Theorem: For constants **a**, A, var (**a**'**x**)=**a**'Σ**a**, var(A**x**+**b**)=AΣA', C(**x**,**x**)=Σ, C(**x**,**y**)=C(**y**,**x**)'

Theorem: If **x**, **y** are independent RVs, then C(**x**,**y**)=0, but not conversely.

Theorem: Let **x**,**y** have same dimension, then var(**x**+**y**)=var(**x**)+var(**y**)+C(**x**,**y**)+C(**y**,**x**)

Normal Random Vectors

The Central Limit Theorem says that if a focal random variable x consists of the sum of many other independent random variables, then the focal random variable will asymptotically have a distribution that is basically of the form e^{-x^2} , which we call "normal" because it is so common.

Normal random variable has pdf $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)\frac{1}{\sigma^2}(x-\mu)/2}$

Denote **x** p×1 normal random variable with pdf

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

where **μ** is the mean vector and Σ is the covariance matrix: $\mathbf{x} \sim N_p(\mu,\Sigma)$.

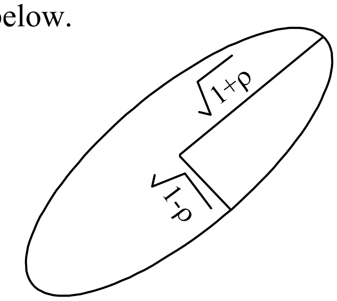
Bivariate Normal $f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2}} e^{-\begin{bmatrix} x_1-\mu_1 \\ x2-\mu_2 \end{bmatrix}' \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1-\mu_1 \\ x2-\mu_2 \end{bmatrix}/2}$ Note

Recall variance σ_{11} is also sometimes written σ_1^2 and by symmetry $\sigma_{12}=\sigma_{21}$. The correlation is $\rho_{12}=\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}} = \sigma_{12}/(\sigma_1\sigma_2)$.

Theorem: Eigenvalue of Σ^{-1} is reciprocal of eigenvalue of Σ and eigenvectors are identical.
Proof: Let $\Sigma^{-1}x=\lambda x$. Then $x=\lambda\Sigma x$ or $\Sigma x=(1/\lambda)x$.

Contour of constant probability is ellipsoid $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)=c^2$ for some c. This is an ellipse centered at μ and with axis that point in the directions of the eigenvectors of Σ with length $c\sqrt{\lambda_i}$, that is the axes are $\pm c\sqrt{\lambda_i}\,\mathbf{e}_i$ where λ_i and $\mathbf{e}_i$ are the eigenvalues and eigenvectors of the covariance matrix Σ.

Suppose that $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, then eigenvalues are defined by $(1-\lambda)^2-\rho^2=0$ or $\lambda=1\pm\rho$. The eigenvectors are values of (x_1,x_2) such that $-\pm\rho x_1+\rho x_2=0$ and $x_1^2+x_2^2=1$; these are $(1/\sqrt{2}, \pm 1/\sqrt{2})$. If the correlation ρ is positive, then the eigenvector $(1/\sqrt{2}, 1/\sqrt{2})$ is stretched to a length greater than 1, $\sqrt{1+\rho}$, while the eigenvector $(1/\sqrt{2}, -1/\sqrt{2})$ is shrunk to a length less than 1, $\sqrt{1-\rho}$. See the figure below.



Theorems: The moment generating function (mgf) for multivariate normal is

$$\phi_{\mathbf{x}}(\mathbf{t}) = E\left[e^{\mathbf{t}'\mathbf{x}}\right] = e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}.$$

$\mathbf{x}\sim N(\mu,\Sigma) \Rightarrow x_i \sim N(\mu_i,\sigma_{ii})$

$y\equiv\mathbf{a}'\mathbf{x} \Rightarrow y\sim N(\mathbf{a}'\mu,\mathbf{a}'\Sigma\mathbf{a})$ and mgf $\phi_y(\tau)=e^{\tau a'\mu+\frac{1}{2}\tau^2 a'\Sigma a}$

$x_1|x_2 \sim N(\mu_1+\sigma_{12}/\sigma_{22}(x_2-\mu_2), \sigma_{11}-\sigma_{12}^2/\sigma_{22})$

Theorem $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) \sim \chi^2_p$

Proof: (note:chi-sq is the sum of the squares of independent normals). Using spectral decomposition, $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)=(\mathbf{x}-\mu)'(P\Lambda P')^{-1}(\mathbf{x}-\mu)=(\mathbf{x}-\mu)'P\Lambda^{-1/2}\Lambda^{-1/2}P'(\mathbf{x}-\mu)$. From above $\Lambda^{-1/2}P'(\mathbf{x}-\mu) \sim N(0, \Lambda^{-1/2}P'\Sigma P\Lambda^{-1/2})=N(0, \Lambda^{-1/2}\Lambda\Lambda^{-1/2})=N(0,I)$, so quadratic form is the sum of independent squared normal random variables. QED

Normal data matrix $X = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix}$ where $\mathbf{x}_i$ is iid $N_p(\mu,\Sigma)$. This is a n×p matrix of random variables.

Each row is independent of other rows and identically distributed.

$$P(X) = (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\left(-\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})/2\right)$$

$$= (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\left[-\mathrm{tr}\left(\Sigma^{-1}\left(\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right)/2\right)\right]$$

Note: $x'Ax=\mathrm{tr}(x'Ax)=\mathrm{tr}(Axx')$

Aside: On Union Intersection Tests

$\mathbf{x}\sim N(\mu,I) \Rightarrow y=\mathbf{a}'\mathbf{x}\sim N(\mathbf{a}'\mu, a'a)$

$H_o{:}\,\mu=\mathbf{0} \Leftrightarrow y_a=\mathbf{a}'\mathbf{x}\sim N(\mathbf{0}, a'a)$ for all $\mathbf{a}$.

H_{oa}: $\mathbf{a}'\mu=0$

$H_o = \cap H_{oa}$ note: if you find one **a** that violates H_{oa} then H_o cannot be true.
Let's test H_{oa} using $z_a \equiv y_a/\sqrt{a'a}$. The rejection region is $R_a=\{z_a | z_a^2>c^2\}$. What about H_o?
$R=\cup R_a$. So H_o is accepted if and only if $z_a^2<c^2$ for all a. The worst case scenario is $\max_a z_a^2$.
So, if $\max_a z_a^2<c^2$ then this will be true for all a.
Suppose that we have independent draws of a random vector. Let $\mathbf{x}_j$ be the jth draw. Define $y=\mathbf{a}'\mathbf{x}_j$. Then we know that $\bar{y}=a'\bar{x}$ and $s_y^2=a'Sa$. Compute $t^2=\dfrac{n(a'(\bar{x}-\mu))^2}{a'Sa}$. Following the Union Intersection test procedure we would like to find the value of a that is the worst case scenario. The Cauchy-Schwartz inequality helps here: $(x'y)^2 \le (x'x)(y'y)$ (this is a consequence of $x'y=\|x\|\,\|y\|\cos\theta$): $(a'(\bar{x}-\mu))^2 \le (a'Sa)\,((\bar{x}-\mu)'S^{-1}(\bar{x}-\mu))$ and can only "=" if $a=S^{-1}(\bar{x}-\mu)$.
Taking this worst case scenario, the $\max_a t^2=n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)$.

Theorem: The interval $a'\bar{x} \pm \sqrt{\dfrac{p(n-1)}{n-p}\dfrac{a'Sa}{n}F_{p,n-p}(\alpha)}$ will contain $a'\mu$ a fraction $1-\alpha$% of the time, simultaneously for all possible a.

Comparison of Traditional and Simultaneous Confidence Intervals

Suppose that you had $H_{oi}:\mu_i=0$ for i=1,2,…,p. If you ignore the fact that there are several simultaneous test, you would do this one variable at a time, computing confidence intervals:

$$\bar{x}_i \pm \sqrt{s_{ii}/n}\cdot t_{n-1}(\alpha/2).$$

As we have seen before, the confidence for these as a whole is not $1-\alpha$%, but rather $(1-\alpha)^p$: for 6 variables $0.95^6=0.75$. Hence the rectangular region sketched out by these intervals is really only a 75% confidence region. If you had 13 variables, then this region will capture the truth in all dimensions only 50% the time. We have a false sense of high accuracy.

The above Union Intersection test would sequentially set a'=(0,..0,1,0,..0) where the 1 is in the ith entry and then calculate the intervals

$$\bar{x}_i \pm \sqrt{s_{ii}/n}\cdot\sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}.$$

These simultaneous intervals will be much wider than the above, but we can then say that there is a 95% confidence that all variables will be covered by the combined rectangle. These intervals are the "shadows of the 95% confidence ellipse in a p-dimensional space.

How much wider are these simultaneous intervals? It depends on n and p. As you can see the intervals are much wider, making it very difficult to say with high confidence, "All elements of my theory are true."

		$\sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}$	
n	$t_{n-1}(0.025)$	p=4	p=10
15	2.145	4.14	11.52
25	2.064	3.60	6.39
50	2.010	3.31	5.05
100	1.970	3.19	4.61
∞	1.960	3.08	4.28

Generalization of t-test to T^2-test

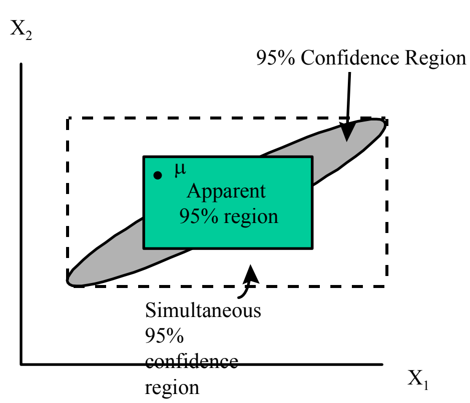
Neither the traditional nor the simultaneous interval tests take into account that the variables may be correlated with one another. The 95% confidence region should not be a rectangle, but rather an ellipse. How should we handle this? This is not that complicated. In the single normal variable case, we test using $t \equiv \frac{\bar{x}-\mu}{s/\sqrt{n}}$. When we have several variables that we want to combined without having –'s canceling +'s, we use the Hotelling T^2-distribution of the variable $t^2 \equiv \frac{(\bar{x}-\mu)^2}{s^2/n}$. For p-variate normal vector case, the equivalent statistic is

$$T^2_{p,n-1} \equiv \frac{(\bar{\mathbf{x}}-\boldsymbol{\mu})' S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})}{1/n} = n(\bar{\mathbf{x}}-\boldsymbol{\mu})' S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}).$$

This statistic has normals $^{\text{squared}}$ on the top (the x'x terms) and normals $^{\text{squared}}$ in the bottom (since S is composed from squared normals). That is, it is the ratio of χ^2s and hence has the F-distribution:

$$T^2_{p,n-1} \sim \frac{(n-1)p}{n-p} F_{p,n-p}.$$

Thus if we had a null hypothesis that the p-variate variable **x** had mean μ, then we would construct the above T^2 statistic and see if it exceeded the critical value found in an F-distribution table. This will tells us whether our theoretical value **μ** is covered by the confidence ellipsoid 1-α% of the time in repeated samples. The 1-α% confidence ellipsoid has axes determined by $n(\bar{\mathbf{x}}-\boldsymbol{\mu})' S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \le c^2 = \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$. That is they are determined by starting at $\bar{\mathbf{x}}$ and going $\pm\sqrt{\lambda_i}\, c/\sqrt{n} = \pm\sqrt{\lambda_i}\sqrt{\frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha)}$ units along the eigenvectors $\mathbf{e}_i$. This is better than doing p separate t-tests of each variable, since it uses <u>all</u> the information in S, including the fact that some variables are highly correlated.



In summary, do not claim when you study p variables and all of them fit your theory that you are 95% confident in your theory. Apparent confidence is not real confidence. On the other hand, even if one-at-a-time you cannot reject the null, you still may be able to with 95% confidence state, "There are some elements of this theory that must be true, I just cannot tell you which ones." In the above graph, the true μ is inside the apparent 95% confidence interval, so you apparently cannot reject any element μ_i, but μ is outside the 95% confidence ellipsoid, so you reject μ as a whole.