Lecture Notes for Theory of Linear Models

- **Distribution of Quadratic Forms (Sum Squares)**

Longhai Li

Department of Mathematics and Statistics
University of Saskatchewan

Quadratic forms

$y = (y_1, \ldots, y_n)'$

$\|y\|^2 = \sum_{i=1}^{n} y_i^2 = y' \cdot I_n y$

$\|Py\|^2 = (Py)' \cdot Py = y' P' P y$
$= y' P y$

where P is a proj matrix

$y' A y$, A may be general.

$E(y^2) = V(y) + u^2, \quad u = E(y)$

Mean of Quadratic Form (without normality assumption)

Theorem 5.2a. If $\mathbf{y}$ is a random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and if $\mathbf{A}$ is a ~~symmetric~~ matrix of constants, then

$$E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}. \tag{5.4}$$

$\Sigma = E\left((y-u)(y-u)'\right)$

PROOF. By (3.25), $\Sigma = E(\mathbf{y}\mathbf{y}') - \boldsymbol{\mu}\boldsymbol{\mu}'$, which can be written as

$E\left(yy' - 2yu' + uu'\right)$

$$E(\mathbf{y}\mathbf{y}') = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}'. \tag{5.5}$$

$= E(yy') - uu'$

Since $\mathbf{y}'\mathbf{A}\mathbf{y}$ is a scalar, it is equal to its trace. We thus have

$\text{tr}(AB) = \text{tr}(BA)$

$$\begin{aligned}
E(\mathbf{y}'\mathbf{A}\mathbf{y}) &= E[\text{tr}(\mathbf{y}'\mathbf{A}\mathbf{y})] \\
&= E[\text{tr}(\mathbf{A}\mathbf{y}\mathbf{y}')] && \text{[by (2.87)]} \\
&= \text{tr}[E(\mathbf{A}\mathbf{y}\mathbf{y}')] && \text{[by (3.5)]} \\
&= \text{tr}[\mathbf{A}E(\mathbf{y}\mathbf{y}')] && \text{[by (3.40)]} \\
&= \text{tr}[\mathbf{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')] && \text{[by (5.8)]} \\
&= \text{tr}[\mathbf{A}\boldsymbol{\Sigma} + A\boldsymbol{\mu}\boldsymbol{\mu}'] && \text{[by (2.15)]} \\
&= \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \text{tr}(\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}) && \text{[by (2.86)]} \\
&= \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

$E(\text{tr}(X)) = \text{tr}(E(X))$

Another proof:

$A = (a_{ij})_{n\times n}$

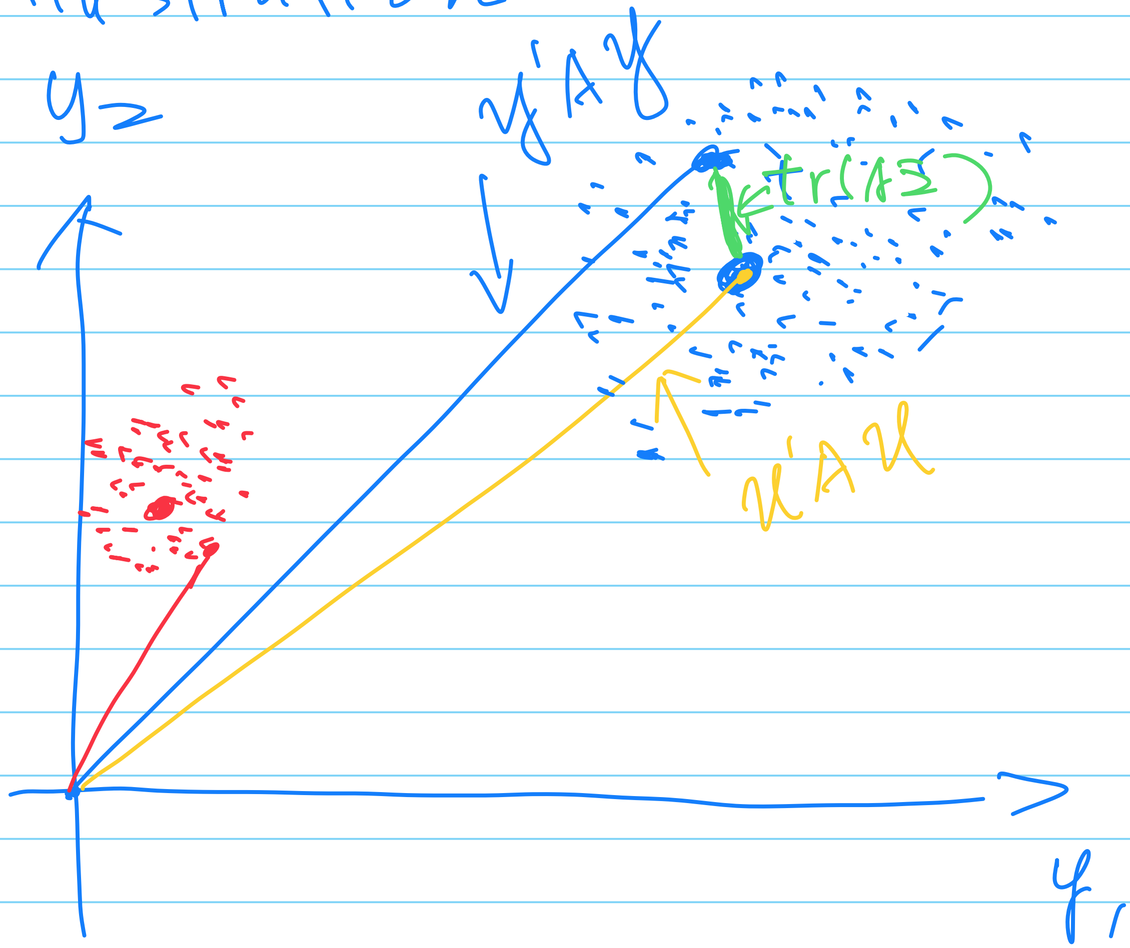
$y'Ay = \sum_i \sum_j y_i y_j a_{ij}$

$E(y_i y_j) = \sigma_{ij} + u_i u_j$

$E(y'Ay) = \sum_i \sum_j \sigma_{ij} a_{ij} + \sum_i \sum_j a_{ij} u_i u_j$

$= \text{tr}(A\Sigma) + u'AU$

Illustration



$A = I_2$, $y'y = \|y\|^2$

Example:

Let x be a random vector with

$E(x) = \mu = (\mu_1, \cdots, \mu_n)'$, $Var(x) = \sigma^2 I_n$

$x = (x_1, \cdots, x_n)'$, $E(x_i) = \mu_i$, $V(x_i) = \sigma^2$

$cor(x_i, x_j) = 0$

Using Thm 5.2a

$$E(\|x\|^2) = E(x' I_n x)$$

$$= tr(I_n \cdot \sigma^2 I_n) + \mu' \cdot I_n \cdot \mu$$

$$= \sigma^2 \cdot n + \sum_{i=1}^{n} \mu_i^2$$

$$E\left(\sum_{i=1}^{n} x_i^2\right) = \sum_{i=1}^{n} E(x_i^2)$$

$$= \sum_{i=1}^{n} (\mu_i^2 + \sigma^2)$$

$$= \sum_{i=1}^{n} \mu_i^2 + n\sigma^2$$

A direct approach:

Let $x_i = z_i + u_i$, where $E(z_i) = 0$, $V(z_i) = \sigma^2$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

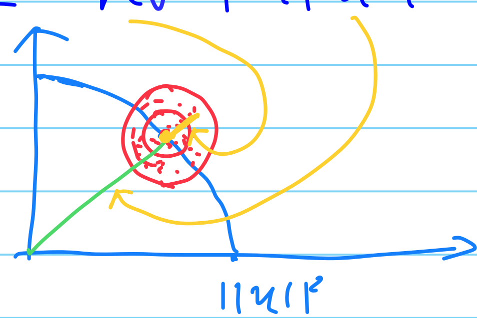
$$\|x\|^2 = \langle z+u, z+u \rangle$$

$$= \|z\|^2 + 2\langle z, u \rangle + \|u\|^2$$

$$E(\|x\|^2) = E(\|z\|^2) + 2E(u'z) + \|u\|^2$$

$$= n\sigma^2 + 2 \cdot u' \cdot \cancel{E(z)}^{0} + \|u\|^2$$

$$= n\sigma^2 + \|u\|^2$$



Example:

$\mu j_n = (\mu, \cdots, \mu)'$

Suppose $E(x) = \mu j_n$, $Var(x) = \sigma^2 I_n$,
μ is a scalar.
x may not follow $N(\mu j_n, \sigma^2 I_n)$

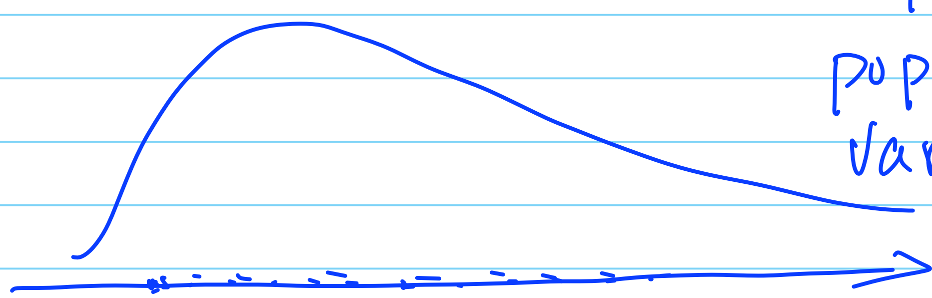
In non-matrix notation

$x = (x_1, \cdots, x_n)'$

$x_1, \cdots, x_n$ are uncorrelated

and $E(x_i) = \mu$, $V(x_i) = \sigma^2$

population variance

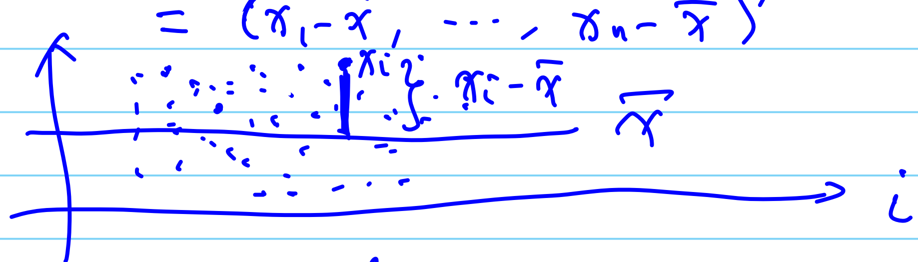


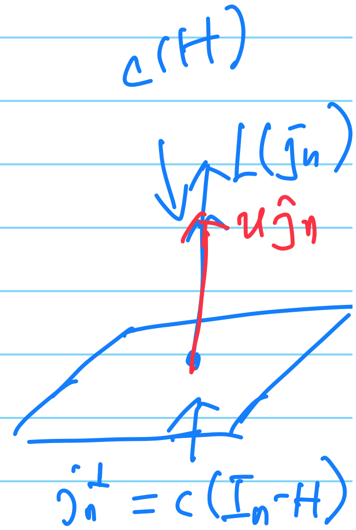
Let $H = \frac{1}{n} \hat{j}_n \hat{j}_n'$. $\quad H x = \bar{x} \cdot \hat{j}_n$

H is a projection matrix onto $L(\hat{j}_n)$

$$(I_n - H)x = x - \text{proj}(x \mid \hat{j}_n) = x - \bar{x} \cdot \hat{j}_n$$

$$= (x_1 - \bar{x}, \dots, x_n - \bar{x})'$$





$$\|(I_n - H)x\|^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \equiv (n-1) S_x^2$$

$$E((n-1)S_x^2) = E(\|(I_n - H)x\|^2)$$

$$= E(x'(I_n - H)x)$$

$$= \text{tr}((I_n - H)\sigma^2 \cdot I_n) + (u \cdot \hat{j}_n)'(I_n - H)u\hat{j}_n$$

$$= \sigma^2 \, \text{tr}(I_n - H) + u^2 \cdot \hat{j}_n'(I_n - H)\hat{j}_n$$

$\uparrow$ 0

$$= \sigma^2 \cdot (n-1)$$

since $\hat{j}_n \perp c(H)$

$$\text{tr}(H) = \text{rank}(H) = 1$$

$$\text{tr}(I_n - H) = \text{rank}(I_n - H) = n - 1$$

In words, sample variance is an unbiased estimate of population variance σ^2.

$$Hx = \text{proj}(x \mid j_n)$$

$$= \frac{1}{n} j_n j_n' x$$

$$= \begin{bmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix}$$

$$E(\|Hx\|^2) = tr(H \cdot \sigma^2 I_n) + \|\mu j_n\|^2$$

$$= \sigma^2 \cdot + n\mu^2$$

$$\underbrace{\longrightarrow}_{n\mu^2+\sigma^2} Hx \underbrace{\longrightarrow}_{(n-1)\sigma^2} x \longrightarrow$$

$\leftarrow$ mean of SS

usig $E(\|Hx\|^2)$ to find $V(\bar{x})$

$$\|Hx\|^2 = n\bar{x}^2$$

$$E(\bar{x}^2) = \frac{n\mu^2 + \sigma^2}{n} = \mu^2 + \frac{\sigma^2}{n}$$

$$V(\bar{x}) = E(\bar{x}^2) - (E(\bar{x}))^2$$

$$= \mu^2 + \frac{\sigma^2}{n} - \mu^2$$

$$= \frac{\sigma^2}{n}$$

Chi-square Distribution: Let $x_1, \ldots, x_n$ be independent normal random variables with means $\mu_1, \ldots, \mu_n$ and common variance 1. Then

$$y = x_1^2 + \cdots + x_n^2 = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2, \quad \text{where } \mathbf{x} = (x_1, \ldots, x_n)^T$$

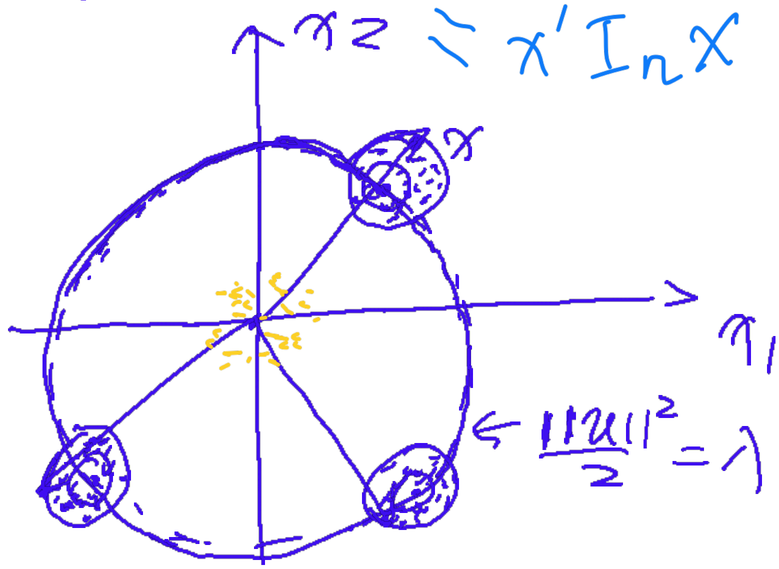
is said to have a **noncentral chi-square distribution** with n **degrees of freedom** and **noncentrality parameter** $\lambda = \frac{1}{2}\sum_{i=1}^n \mu_i^2$. We denote this as $y \sim \chi^2(n, \lambda)$.

In matrix form,

Let $x \sim N(u, I_n)$, $u = (u_1, \ldots, u_n)'$

$$y = \|x\|^2 = x'x \sim \chi^2\left(n, \frac{\|u\|^2}{2}\right)$$

$\doteq x' I_n x$



The distribution of $\|x\|^2$ is determined by $\|u\|^2$, rather than the specific u.

PDF

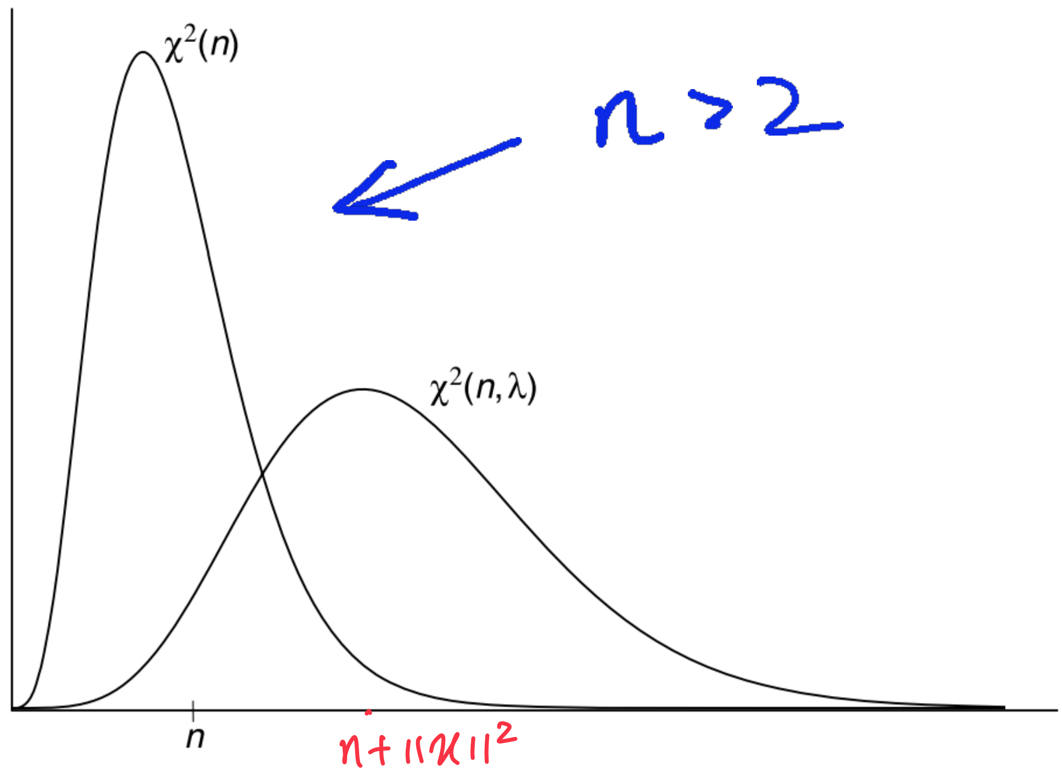


Figure 5.1 Central and noncentral chi-square densities.

Mean, Variance and MGF

$\lambda = \frac{\|\mu\|^2}{2}$

Theorem: Let $Y \sim \chi^2(n, \lambda)$. Then

i. $E(Y) = n + 2\lambda$; $= n + \|\mu\|^2$
ii. $\text{var}(Y) = 2n + 8\lambda$; and
iii. the m.g.f. of Y is

$$m_Y(t) = \frac{\exp[-\lambda\{1 - 1/(1-2t)\}]}{(1-2t)^{n/2}}.$$

Pf:

i) $Y = \|x\|^2$, with $x \sim N(\mu, I_n)$

$Y = \|\mu + z\|^2$, where $z \sim N(0, I_n)$

$= \|\mu\|^2 + \|z\|^2 + 2\mu' z$

$\|z\|^2 \sim \chi^2_n$ (central).

$E(\|z\|^2) = n$

$E(Y) = \mu^2 + n + 2 \cdot \mu' \cdot 0 = \mu^2 + n$

ii) with M.G.F.

iii) a special case of Thm 5.2b

Additivity

Theorem 5.3c. If $v_1, v_2, \ldots, v_k$ are independently distributed as $\chi^2(n_i, \lambda_i)$, then

$$\sum_{i=1}^{k} v_i \text{ is distributed as } \chi^2\left(\sum_{i=1}^{k} n_i, \sum_{i=1}^{k} \lambda_i\right). \tag{5.27}$$

□

pf: using M.G.F.

$$M_{\sum_{i=1}^{k} v_i}(t) = \frac{\exp\left(\sum_{i=1}^{k} \lambda_i \left(1 - \frac{1}{1-2t}\right)\right)}{(1-2t)^{\frac{\sum_{i=1}^{k} n_i}{2}}}$$

This is the M.G.F. of $\chi^2(\Sigma n_i, \Sigma \lambda_i)$

pf2:

$v_1 = \|x_1\|^2, \quad x_1 \sim N(\mu_1, I_{n_1})$

$v_2 = \|x_2\|^2, \quad x_2 \sim N(\mu_2, I_{n_2})$

$\vdots$

$v_k = \|x_k\|^2, \quad x_k \sim N(\mu_k, I_{n_k})$

$\sum_{i=1}^{k} v_i = \sum_{i=1}^{k} \|x_i\|^2 = \|x\|^2,$

where $x = (x_1', x_2', \ldots, x_k')'$

$\sim N((\mu_1', \ldots, \mu_k')', I_{n_1+\cdots+n_k})$

MGF of Quadratic Form

Theorem 5.2b. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the moment generating function of $\mathbf{y}'\mathbf{A}\mathbf{y}$ is

$$M_{\mathbf{y}'\mathbf{A}\mathbf{y}}(t) = |\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}|^{-1/2} e^{-\boldsymbol{\mu}'[\mathbf{I}-(\mathbf{I}-2t\mathbf{A}\boldsymbol{\Sigma})^{-1}]\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/2}$$

The distribution of $y'Ay$

$A = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$

Σ is p.d.

Theorem 5.5. Let $\mathbf{y}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, let $\mathbf{A}$ be a symmetric matrix of constants of rank r, and let $\lambda = \frac{1}{2}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$. Then $\mathbf{y}'\mathbf{A}\mathbf{y}$ is $\chi^2(r, \lambda)$, if and only if $\mathbf{A}\boldsymbol{\Sigma}$ is idempotent. ($A\Sigma$ may not be a projection matrix)

Pf: 1) using the M.G.F. of $y'Ay$.
See the textbook. (very complicated)
2) Will be proved in next pages.

Important:

special

Corollary Suppose $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$ and let $\mathbf{P}_V$ be the projection matrix onto a subspace $V \in \mathcal{R}^n$ of dimension $r \leq n$. Then

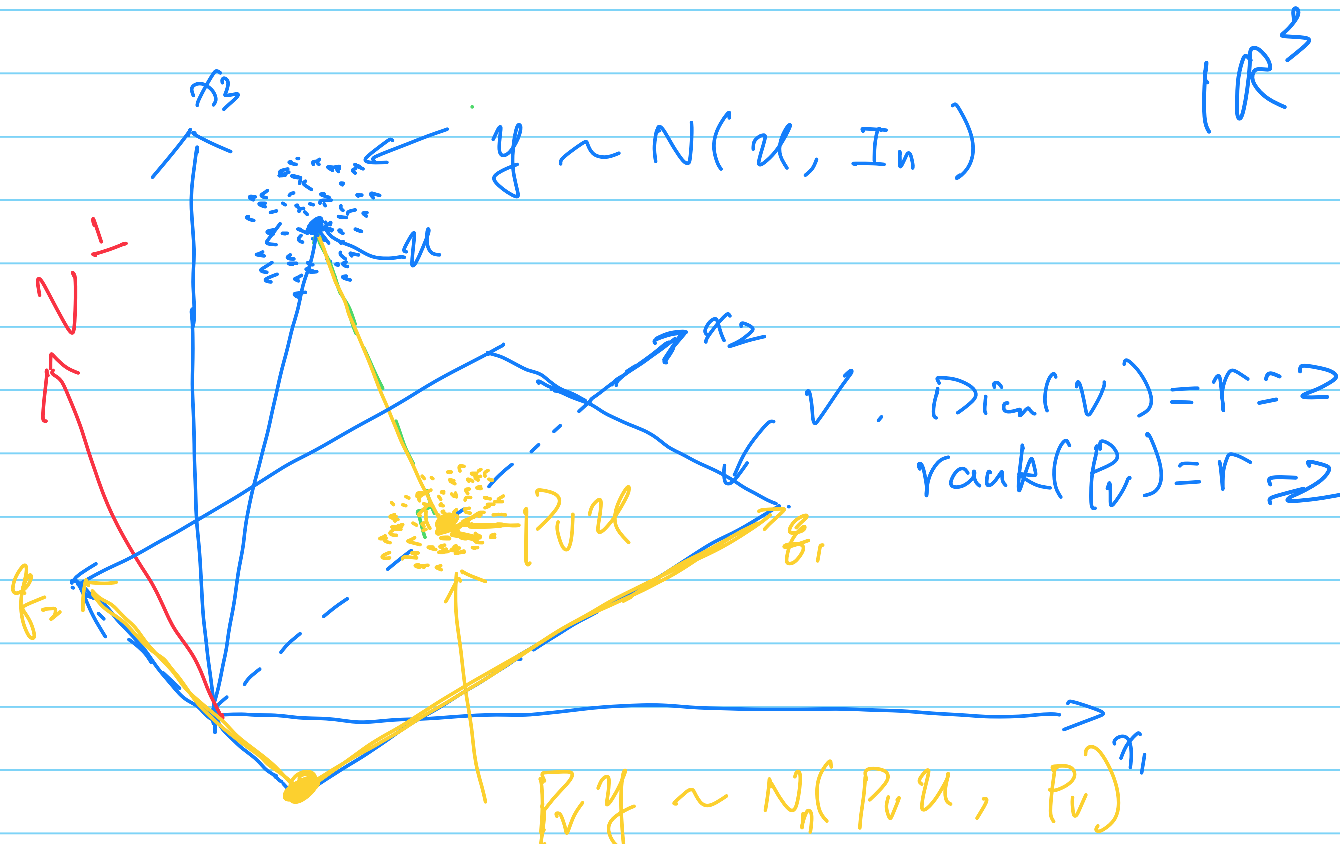
$$\frac{1}{\sigma^2}\mathbf{y}^T\mathbf{P}_V\mathbf{y} = \frac{1}{\sigma^2}\|p(\mathbf{y}|V)\|^2 \sim \chi^2(r, \frac{1}{2\sigma^2}\boldsymbol{\mu}^T\mathbf{P}_V\boldsymbol{\mu}) = \chi^2(r, \frac{1}{2\sigma^2}\|p(\boldsymbol{\mu}|V)\|^2).$$

$$P_V = \underset{n\times n}{Q^*} \cdot \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} (Q^*)'$$

$$P_V y = Q^* \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q^{*\prime} y$$

$$= (Q, Q_2) \cdot \begin{pmatrix} I_r, & 0 \\ 0, & 0 \end{pmatrix} \begin{pmatrix} Q' \\ Q_2' \end{pmatrix} y$$

$$= Q Q' y$$



$$\underset{n\times r}{P_V = Q}\,\underset{r\times n}{Q'}, \text{ where } Q'Q = I_r$$

$$\underset{r\times n}{Q'}\, y \sim N(Q'\mu, I_r)$$

$Q'y \in \mathbb{R}^r$

$Q'y$ is the coordinates of the projection of y onto $V = L(q_1, \dots, q_r)$

Pf of the corollary:

suppose $\sigma^2=1$ (later generalized to $\sigma^2\neq 1$)

$$y' P_V y = y' P_V' P_V y$$

$$= \| P_V y \|^2$$

By the formular for linear transformation of multivariate normal

$$P_V y \sim N_n(P_V \mu, P_V I_n \cdot P_V')$$

$$= N_n(P_V \mu, P_V)$$

$\text{rank}(P_V) = r < n$, $P_V y \in \mathbb{R}^n$

The distribution of $P_V y$ is degenerated (confined to the space V).

We will express $\|P_V y\|^2$ with a non-degenerated random vector in $\mathbb{R}^r$

$$P_V = Q \cdot Q'$$
$$= (q_1, \cdots, q_r) \cdot \begin{bmatrix} q_1' \\ q_2' \\ \vdots \\ q_r' \end{bmatrix},$$

with $Q'Q = I_r$.

$$P_V y = Q \cdot Q' y$$
$$= (q_1, \cdots, q_r) \cdot \begin{bmatrix} q_1' y \\ q_2' y \\ \vdots \\ q_r' y \end{bmatrix}$$

$Q'y \in \mathbb{R}^r$, $P_V y \in \mathbb{R}^n$

$$\| P_V y \|^2 = y' Q Q' \cdot Q \cdot Q' y$$
$$= \| Q' y \|^2, \text{ for all } y \in \mathbb{R}^n$$

$Q'y$ is the coordinates of the projection of y onto $V = L(q_1, \cdots, q_r)$

$$Q'y \sim N_r(Q'\mu,\ Q'I_nQ = I_r)$$

By the definition of χ^2,

$$\|Q'y\|^2 \sim \chi^2\left(r,\ \tfrac{1}{2}\|Q'\mu\|^2\right)$$

$$= \chi^2\left(r,\ \tfrac{1}{2}\|P_V\mu\|^2\right)$$

So $\|P_V y\|^2 \sim \chi^2\left(r,\ \tfrac{1}{2}\|P_V\mu\|^2\right)$

Note:

$\|Q'\mu\|^2 = \|P_V\mu\|^2$ because

$\|Q'y\|^2 = \|P_V y\|^2$ for all $y \in \mathbb{R}^n$

Now, if $\sigma^2 \neq 1$, suppose

$$y \sim N(u, \sigma^2 I_n)$$

$$\Rightarrow \frac{y}{\sigma} \sim N\left(\frac{u}{\sigma}, I_n\right)$$

Applying previous result for $\sigma^2 = 1$

$$\left\| P_V\left(\frac{y}{\sigma}\right)\right\|^2 \sim \chi^2\left(r, \frac{1}{2}\left\| P_V\left(\frac{u}{\sigma}\right)\right\|^2\right)$$

$$\frac{\| P_V y \|^2}{\sigma^2} \sim \chi^2\left(r, \frac{1}{2}\frac{\| P_V u \|^2}{\sigma^2}\right)$$

Note: this doesn't say that

$$\| P_V y \|^2 \sim \chi^2\left(r, \frac{1}{2}\| P_V u \|^2\right)$$

Note: $$E\left(\frac{\| P_V y \|^2}{\sigma^2}\right) = r + \frac{\| P_V u \|^2}{\sigma^2}$$

$$E\left(\| P_V y \|^2\right) = r \cdot \sigma^2 + \| P_V u \|^2$$

Alternatively, $E\left(\| P_V y \|^2\right)$

$$= tr\left(P_V \cdot \sigma^2 I\right) + \| P_V u \|^2$$

$$= r \cdot \sigma^2 + \| P_V u \|^2$$

Lemma: not p.s.d.

Σ is p.d. $\left(\exists\ \Sigma^{\frac{1}{2}} \text{ s.t. } \Sigma = \Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}}\right)$

$A\Sigma$ is idempotent $\Leftrightarrow$ $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ is idemp.

Pf: $(\Rightarrow)$ $A\Sigma A\Sigma = A\Sigma$ (given)

$$\left(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}\right)\left(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}\right)$$

$$= \Sigma^{\frac{1}{2}} \underline{A \Sigma A \Sigma}\, \Sigma^{-\frac{1}{2}}$$

$$= \Sigma^{\frac{1}{2}} A \Sigma \cdot \Sigma^{-\frac{1}{2}}$$

$$= \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$$

"$\Leftarrow$" $\left(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}\right)\left(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}\right) = \Sigma^{\frac{1}{2}} A \Sigma A \Sigma^{\frac{1}{2}}$

$$= \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$$

$\Rightarrow$ $A \Sigma A = A$

$\Rightarrow$ $A \Sigma A \Sigma = A \Sigma$

Thm 5.5 (A) (only one direction)

Let A be a symmetric matrix with $\text{rank}(A) = r$.
$y \sim N(\mu, \Sigma)$, Σ^{-1} exists. If $A\Sigma$ is idempotent,

$$y'Ay \sim \chi^2(r, \tfrac{1}{2}\mu' A \mu)$$

Pf:

Let $y^* = \Sigma^{-\frac{1}{2}} y$. $y^* \sim N_n(\Sigma^{-\frac{1}{2}}\mu, I_n)$

$y'Ay = y'\Sigma^{-\frac{1}{2}}(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})\Sigma^{-\frac{1}{2}} y = y^{*\prime} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} y^*$

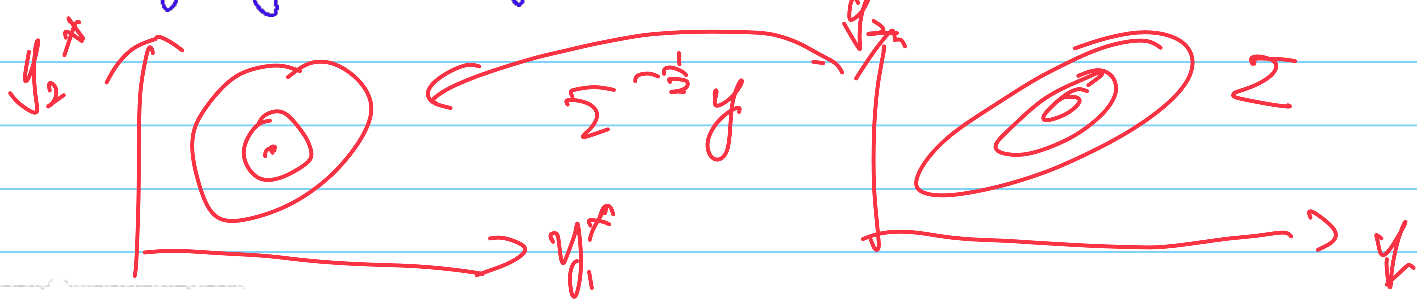
Let $P_V = \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$. $y'Ay = y^{*\prime} P_V y^* = \|P_V y^*\|^2$

$A\Sigma$ is idempotent $\Rightarrow P_V$ is idempotent.
A is symmetric $\Rightarrow P_V = P_V'$

P_V is a proj. matrix with rank r

Applying Cor with $\sigma^2 = 1$,

$y'Ay = \|P_V y^*\|^2 \sim \chi^2(r, \tfrac{1}{2}\|P_V \Sigma^{-\frac{1}{2}}\mu\|^2)$



$$\|P_V \Sigma^{-\frac{1}{2}} u\|^2 = u' \Sigma^{-\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} u$$

$$= u' A u$$

$$\text{rank}(P_V) = \text{rank}(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})$$

$$= \text{rank}(A), \text{ since } \Sigma^{-\frac{1}{2}} \text{ p.d.}$$

Corallary:

suppose $y \sim N_n(\mu, \Sigma)$, $y \in \mathbb{R}^n$

Then,

$$(y-\mu_0)'\Sigma^{-1}(y-\mu_0)$$

$$\sim \chi^2\left(n, \tfrac{1}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)\right)$$

Pf: Let $A = \Sigma^{-1}$, $A\Sigma = I_n$

$$y-\mu_0 \sim N(\mu-\mu_0, \Sigma)$$

We can also prove the corallary directly:

$$\Sigma^{-\frac{1}{2}}(y-\mu_0) \sim N_n(\Sigma^{-\frac{1}{2}}(\mu-\mu_0), I_n)$$

Therefore, by the definition of χ^2:

$$(y-\mu_0)'\Sigma^{-1}(y-\mu_0) = \|\Sigma^{-\frac{1}{2}}(y-\mu_0)\|^2$$

$$\sim \chi^2\left(n, \lambda = \tfrac{1}{2}\|\Sigma^{-\frac{1}{2}}(\mu-\mu_0)\|^2\right)$$

Distributions of a projection and its Sum Square

Theorem: Let V be a $k-$dimensional subspace of $\mathcal{R}^n$, and let $\mathbf{y}$ be a random vector in $\mathcal{R}^n$ with mean $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}$. Then

1. $\mathrm{E}\{p(\mathbf{y}|V)\} = p(\boldsymbol{\mu}|V)$;

   $p(y|V) = P_V \cdot y$

2. if $\mathrm{var}(\mathbf{y}) = \sigma^2 \mathbf{I}_n$ then

   $$\mathrm{var}\{p(\mathbf{y}|V)\} = \sigma^2 \mathbf{P}_V \quad \text{and} \quad \mathrm{E}\left\{\|p(\mathbf{y}|V)\|^2\right\} = \sigma^2 k + \|p(\boldsymbol{\mu}|V)\|^2;$$

   reduced rank

   and

3. if we assume additionally that $\mathbf{y}$ is m'variate normal i.e., $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$, then

   $$p(\mathbf{y}|V) \sim N_n(p(\boldsymbol{\mu}|V), \sigma^2 \mathbf{P}_V),$$

   and

   $$\frac{1}{\sigma^2}\|p(\mathbf{y}|V)\|^2 = \frac{1}{\sigma^2}\mathbf{y}^T \mathbf{P}_V \mathbf{y} \sim \chi^2(k, \frac{1}{2\sigma^2} \underbrace{\boldsymbol{\mu}^T \mathbf{P}_V \boldsymbol{\mu}}_{=\|p(\boldsymbol{\mu}|V)\|^2}).$$

Proof:

1. Since the projection operation is linear, $\mathrm{E}\{p(\mathbf{y}|V)\} = p(\mathrm{E}(\mathbf{y})|V) = p(\boldsymbol{\mu}|V)$.

   $E(P_V y) = P_V \cdot E(y) = P_V \mu$

2. $p(\mathbf{y}|V) = \mathbf{P}_V \mathbf{y}$ so $\mathrm{var}\{p(\mathbf{y}|V)\} = \mathrm{var}(\mathbf{P}_V \mathbf{y}) = \mathbf{P}_V \sigma^2 \mathbf{I}_n \mathbf{P}_V^T = \sigma^2 \mathbf{P}_V$. In addition, $\|p(\mathbf{y}|V)\|^2 = p(\mathbf{y}|V)^T p(\mathbf{y}|V) = (\mathbf{P}_V \mathbf{y})^T \mathbf{P}_V \mathbf{y} = \mathbf{y}^T \mathbf{P}_V \mathbf{y}$. So, $\mathrm{E}(\|p(\mathbf{y}|V)\|^2) = \mathrm{E}(\mathbf{y}^T \mathbf{P}_V \mathbf{y})$ is the expectation of a quadratic form and therefore equals

   $$\begin{aligned}\mathrm{E}(\|p(\mathbf{y}|V)\|^2) = \mathrm{tr}(\sigma^2 \mathbf{P}_V) + \boldsymbol{\mu}^T \mathbf{P}_V \boldsymbol{\mu} &= \sigma^2 \mathrm{tr}(\mathbf{P}_V) + \boldsymbol{\mu}^T \mathbf{P}_V^T \mathbf{P}_V \boldsymbol{\mu} \\ &= \sigma^2 k + \|p(\boldsymbol{\mu}|V)\|^2.\end{aligned}$$

3. The previous Cor

Orthogonal Projections and their Quadratic Forms

Theorem: Let $V_1, \ldots, V_k$ be mutually orthogonal subspaces of $\mathcal{R}^n$ with dimensions $d_1, \ldots, d_k$, respectively, and let $\mathbf{y}$ be a random vector taking values in $\mathcal{R}^n$ which has mean $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}$. Let $\mathbf{P}_i$ be the projection matrix onto V_i so that $\hat{\mathbf{y}}_i = p(\mathbf{y}|V_i) = \mathbf{P}_i\mathbf{y}$ and let $\boldsymbol{\mu}_i = \mathbf{P}_i\boldsymbol{\mu}$, $i = 1, \ldots, n$. Then

1. if $\mathrm{var}(\mathbf{y}) = \sigma^2\mathbf{I}_n$ then $\mathrm{cov}(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) = \mathbf{0}$, for $i \neq j$; and
2. if $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$ then $\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_k$ are independent, with

$$\hat{\mathbf{y}}_i \sim N(\boldsymbol{\mu}_i, \sigma^2\mathbf{P}_i); \quad \leftarrow \text{degenerated!}$$

   and

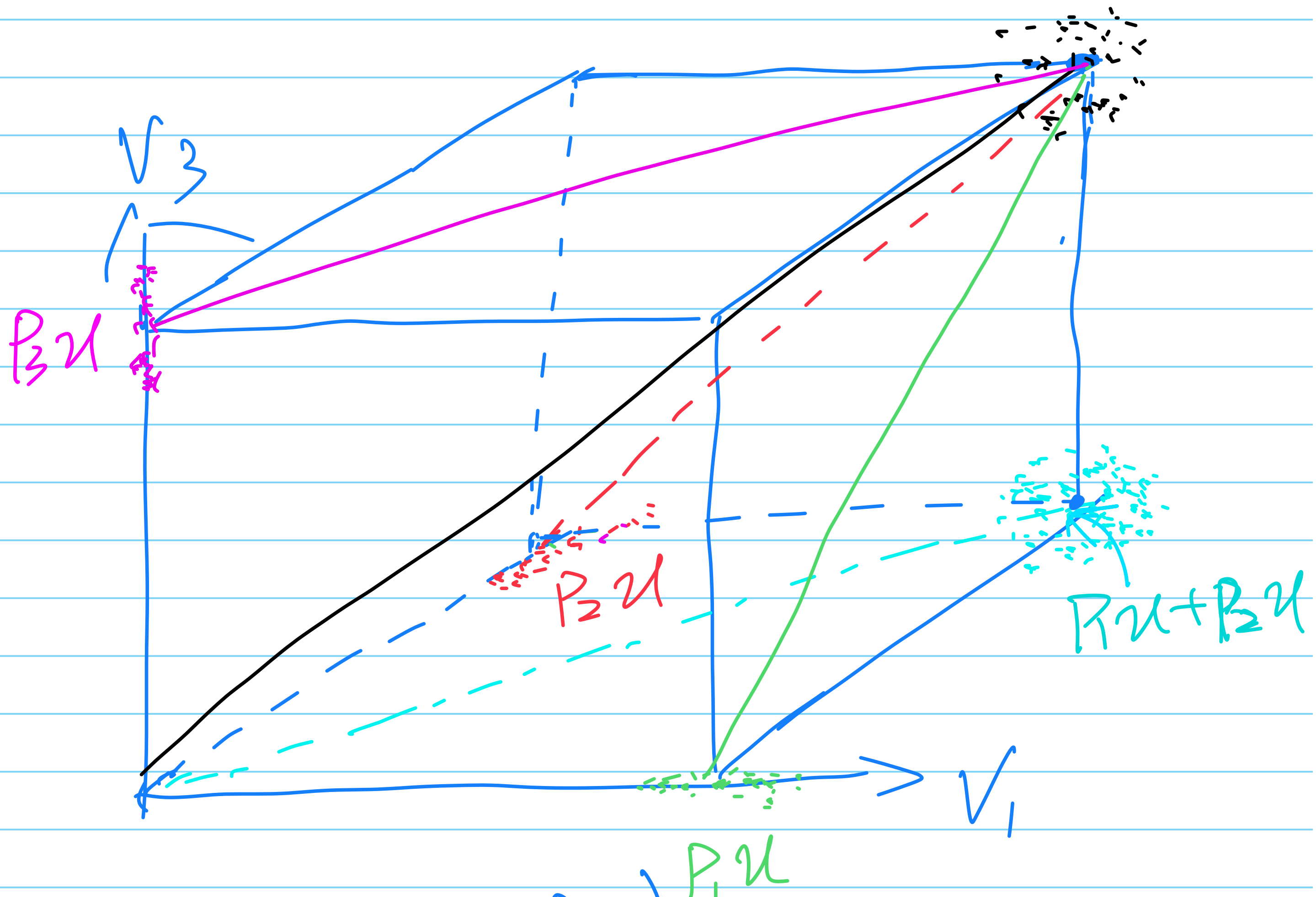
3. if $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$ then $\|\hat{\mathbf{y}}_1\|^2, \ldots, \|\hat{\mathbf{y}}_k\|^2$ are independent, with

$$\frac{1}{\sigma^2}\|\hat{\mathbf{y}}_i\|^2 \sim \chi^2(d_i, \frac{1}{2\sigma^2}\|\boldsymbol{\mu}_i\|^2).$$

Proof: Part 1: For $i \neq j$, $\mathrm{cov}(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) = \mathrm{cov}(\mathbf{P}_i\mathbf{y}, \mathbf{P}_j\mathbf{y}) = \mathbf{P}_i\mathrm{cov}(\mathbf{y}, \mathbf{y})\mathbf{P}_j = \mathbf{P}_i\sigma^2\mathbf{I}\mathbf{P}_j = \sigma^2\mathbf{P}_i\mathbf{P}_j = \mathbf{0}$. (For any $\mathbf{z} \in \mathcal{R}^n$, $\mathbf{P}_i\mathbf{P}_j\mathbf{z} = \mathbf{0} \Rightarrow \mathbf{P}_i\mathbf{P}_j = \mathbf{0}$.)

Part 2: If $\mathbf{y}$ is m'variate normal then $\hat{\mathbf{y}}_i = \mathbf{P}_i\mathbf{y}$, $i = 1, \ldots, k$, are jointly multivariate normal and are therefore independent if and only if $\mathrm{cov}(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) = \mathbf{0}$, $i \neq j$. The mean and variance-covariance of $\hat{\mathbf{y}}_i$ are $\mathrm{E}(\hat{\mathbf{y}}_i) = \mathrm{E}(\mathbf{P}_i\mathbf{y}) = \mathbf{P}_i\boldsymbol{\mu} = \boldsymbol{\mu}_i$ and $\mathrm{var}(\hat{\mathbf{y}}_i) = \mathbf{P}_i\sigma^2\mathbf{I}\mathbf{P}_i^T = \sigma^2\mathbf{P}_i$.

Part 3: If $\hat{\mathbf{y}}_i = \mathbf{P}_i\mathbf{y}$, $i = 1, \ldots, k$, are mutually independent, then any (measurable*) functions $f_i(\hat{\mathbf{y}}_i)$, $i = 1, \ldots, k$, are mutually independent. Thus $\|\hat{\mathbf{y}}_i\|^2$, $i = 1, \ldots, k$, are mutually independent. That $\sigma^{-2}\|\hat{\mathbf{y}}_i\|^2 \sim \chi^2(d_i, \frac{1}{2\sigma^2}\|\boldsymbol{\mu}_i\|^2)$ follows from part 3 of the previous theorem. ∎



$$y \sim N(\mu, \sigma^2 I_n)$$

$$E(\|Py\|^2) = \|P\mu\|^2 + \sigma^2 \operatorname{rank}(P)$$

$$\frac{\|Py\|^2}{\sigma^2} \sim \chi^2\left(\operatorname{rank}(P), \frac{1}{2}\frac{\|P\mu\|^2}{\sigma^2}\right)$$

Independence of Linear and Quadratic Forms under MVN

Theorem: Suppose $\mathbf{B}$ is a $k \times n$ matrix of constants, $\mathbf{A}$ a $n \times n$ symmetric matrix of constants, and $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \Sigma)$. Then $\mathbf{By}$ and $\mathbf{y}^T\mathbf{Ay}$ are independent if and only if $\mathbf{B}\Sigma\mathbf{A} = \mathbf{0}_{k\times n}$.

Proof

Assuming $\mathbf{A}$ is symmetric and idempotent, then we have

$$\mathbf{y}^T\mathbf{Ay} = \mathbf{y}^T\mathbf{A}^T\mathbf{Ay} = \|\mathbf{Ay}\|^2.$$

Now suppose $\mathbf{B}\Sigma\mathbf{A} = \mathbf{0}$. Then $\mathbf{By}$ and $\mathbf{Ay}$ are each normal, with

$$\mathrm{cov}(\mathbf{By}, \mathbf{Ay}) = \mathbf{B}\Sigma\mathbf{A} = \mathbf{0}.$$

Therefore, $\mathbf{By}$ and $\mathbf{Ay}$ are independent of one another. Furthermore, $\mathbf{By}$ is independent of any (measureable) function of $\mathbf{Ay}$, so that $\mathbf{By}$ is independent of $\|\mathbf{Ay}\|^2 = \mathbf{y}^T\mathbf{Ay}$.

Theorem: Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrices of constants. If $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \Sigma)$, then $\mathbf{y}^T\mathbf{Ay}$ and $\mathbf{y}^T\mathbf{By}$ are independent if and only if $\mathbf{A}\Sigma\mathbf{B} = \mathbf{0}$.

$Ay,\ y'Ay \perp By,\ y'By$

Cochran theorem

Theorem: Let $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$, let $\mathbf{A}_i$ be symmetric of rank r_i, $i = 1, \dots, k$, and let $\mathbf{A} = \sum_{i=1}^k \mathbf{A}_i$ with rank r so that $\mathbf{y}^T\mathbf{A}\mathbf{y} = \sum_{i=1}^k \mathbf{y}^T\mathbf{A}_i\mathbf{y}$. Then

1. $\mathbf{y}^T\mathbf{A}_i\mathbf{y}/\sigma^2 \sim \chi^2(r_i, \boldsymbol{\mu}^T\mathbf{A}_i\boldsymbol{\mu}/\{2\sigma^2\})$, $i = 1, \dots, k$; and
2. $\mathbf{y}^T\mathbf{A}_i\mathbf{y}$ and $\mathbf{y}^T\mathbf{A}_j\mathbf{y}$ are independent for all $i \neq j$; and
3. $\mathbf{y}^T\mathbf{A}\mathbf{y}/\sigma^2 \sim \chi^2(r, \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}/\{2\sigma^2\})$;

if and only if any two of the following statements are true:

a. each $\mathbf{A}_i$ is idempotent;
b. $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$ for all $i \neq j$;
c. $\mathbf{A}$ is idempotent;

$A_1 y, \dots, A_k y$ are projections to orthogonal space

or if and only if (c) and (d) are true where (d) is as follows:

d. $r = \sum_{i=1}^k r_i$.

Corollary: Let $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, let $\mathbf{A}_i$ be symmetric of rank r_i, $i = 1, \dots, k$, and suppose that $\mathbf{y}^T\mathbf{y} = \sum_{i=1}^k \mathbf{y}^T\mathbf{A}_i\mathbf{y}$ (i.e., $\sum_{i=1}^k \mathbf{A}_i = \mathbf{I}$). Then

1. each $\mathbf{y}^T\mathbf{A}_i\mathbf{y} \sim \chi^2(r_i, \boldsymbol{\mu}^T\mathbf{A}_i\boldsymbol{\mu}/\{2\sigma^2\})$; and
2. the $\mathbf{y}^T\mathbf{A}_i\mathbf{y}$'s are mutually independent;

$V_1 \oplus V_2 + \cdots \oplus V_k = \mathbb{R}^n$

if and only if any one of the following statements holds:

a. each $\mathbf{A}_i$ is idempotent;
b. $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$ for all $i \neq j$;
c. $n = \sum_{i=1}^k r_i$.

$SS_i = y' A_i y$

SS_1 SS_2 SS_3 SS_k y

$\|y\|^2 = SS_1 + SS_2 + \cdots + SS_k$

F-Distribution: Let

$$U_1 \sim \chi^2(n_1, \lambda), \qquad U_2 \sim \chi^2(n_2) \quad \text{(central)}$$

be independent. Then

$$V = \frac{U_1/n_1}{U_2/n_2}$$

is said to have a noncentral F distribution with noncentrality parameter λ, and n_1 and n_2 degrees of freedom.

t-Distribution: Let

$$W \sim N(\mu, 1), \qquad Y \sim \chi^2(m)$$

be independent random variables. Then

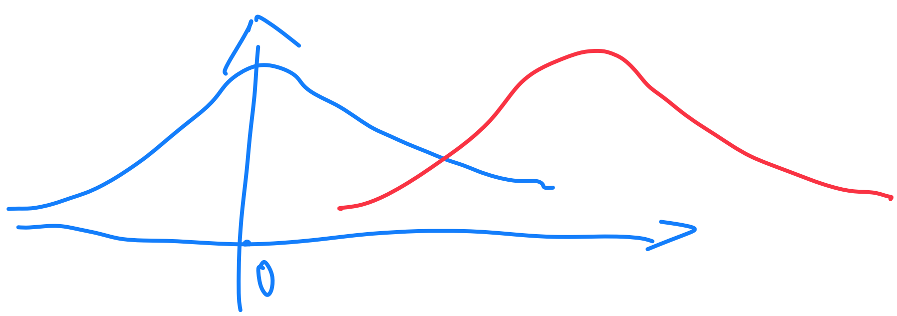
$t(m, \mu)$

$$T = \frac{W}{\sqrt{Y/m}}$$

is said to have a (Student's) t distribution with noncentrality parameter μ and m degrees of freedom.

different from the λ in $\chi^2(n, \lambda)$

$\lambda = \frac{1}{2} \|\mu\|^2$



Example: Student's

Theorem: Let $Y_1, \ldots, Y_n$ be a random sample (i.i.d. r.v.'s) from a $N(\mu, \sigma^2)$ distribution, and let $\bar{Y}, S^2$, and T be defined as above. Then

1. $\bar{Y} \sim N(\mu, \sigma^2/n)$;
2. $V = \frac{S^2(n-1)}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sim \chi^2(n-1)$;
3. $\bar{Y}$ and S^2 are independent; and
4. $T = \frac{\bar{Y}-\mu_0}{S/\sqrt{n}} \sim t(n-1, \lambda)$ where $\lambda = \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$ for any constant μ_0.

Pf: Let $y = (Y_1, \cdots, Y_n)'$

$y \sim N_n(\mu \tilde{j}_n, \sigma^2 I_n)$, Let $\mu_y = \mu \tilde{j}_n$

1) $\bar{Y} = \frac{1}{n} \tilde{j}_n' y \sim N(\mu, \frac{\sigma^2}{n})$.

Note: $\frac{1}{n} \tilde{j}_n' \cdot \mu \tilde{j}_n = \mu$

$\frac{1}{n} \tilde{j}_n' \sigma^2 I_n \tilde{j}_n \cdot \frac{1}{n} = \frac{\sigma^2}{n}$

2) $S^2 \cdot (n-1) = \sum (Y_i - \bar{Y})^2$

$= \| (I_n - P_{\tilde{j}_n}) y \|^2$

$= \| P_V y \|^2$

$$\begin{bmatrix} \bar{Y} \\ \vdots \\ \bar{Y} \end{bmatrix} = P_{\tilde{j}_n} y$$

where $P_V = I_n - \frac{1}{n} \tilde{j}_n \tilde{j}_n'$, $\text{rank}(P_V) = n-1$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\|P_V y\|^2}{\sigma^2} = \frac{y' P_V y}{\sigma^2}$$

$$\sim \chi^2\left(n-1, \frac{1}{2} \frac{\|P_V \mu_y\|^2}{\sigma^2}\right)$$

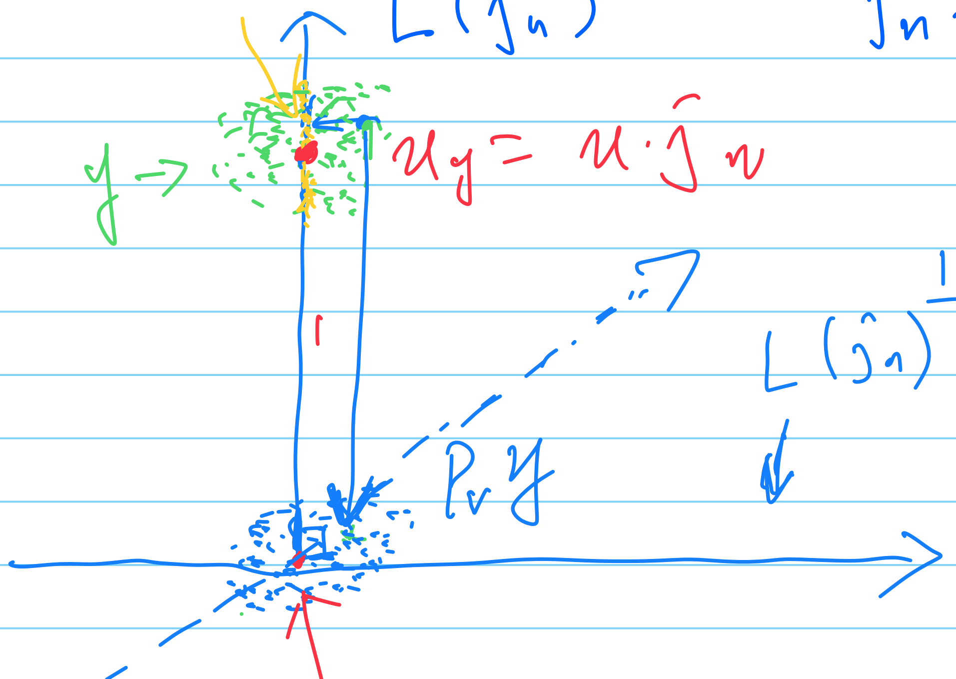
$\|P_V \mu_y\|^2 = \| P_V \cdot \tilde{j}_n \cdot \mu \|^2 = 0$

since $\tilde{j}_n \perp P_V$, $\quad P_V \tilde{j}_n = 0$

so $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1, 0)$. (Central)

$$P_{\hat{j}_n} y = \left(\frac{1}{n}\hat{j}_n \hat{j}_n^{T}\right) y$$

$$\hat{j}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$



$$P_V u_y = 0$$

(3) $\bar{Y} = \frac{1}{n} J_n' y$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{y' P_V y}{\sigma^2}$$

$$\frac{1}{n} J_n' \cdot P_V = 0$$

so $\bar{Y}$ indep. $\frac{(n-1)s^2}{\sigma^2}$

so $\bar{Y}$ indep s^2

Alternative proof:

$$P_{J_n} y = [\bar{Y}, \dots, \bar{Y}]' \perp P_V y$$

$$\Downarrow \qquad\qquad\qquad \Downarrow$$

$$\bar{Y} = [1, 0, \dots, 0]' P_{J_n} y \perp y' P_V y$$

(4) $t = \dfrac{\bar{Y} - u_0}{S/\sqrt{n}} = \dfrac{\sqrt{n}(\bar{Y} - u_0)/\sigma}{\sqrt{\dfrac{(n-1)S^2/\sigma^2}{n-1}}}$

$$\frac{\sqrt{n}(\bar{Y} - u_0)}{\sigma} \sim N\left(\frac{\sqrt{n}(u - u_0)}{\sigma}, 1\right),$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1, 0)$$