Prof. H. Bensmail
Stat-664: Adv Stat Infer

Spring Semester 2006

Chapter 1: Review

1.1 Univariate discrete distribution:

X: random
S: sample space of X (number of values)
Ω : Space of possible outcome
P_X : probability on S

G $\subseteq$ S, a subset of S

$$\begin{aligned} P_X(G) &= P(\omega \in \Omega,\ X \in G) \\ &= P(X \in G) \end{aligned}$$

P_X : is the distribution of X defined on G.
$P \in [0,1] \qquad P \geq 0 \qquad P \nleq 0$

Probability function (p.d.f)

$$P(x) = \left[\begin{array}{cc} P(X = x) & \text{if } x \in S \\ 0 & \text{Otherwise} \end{array} \right]$$

x	2	1	0
$P(X = x)$	1/4	1/2	1/4

$$P(X = 2) = p(2) = 1/4$$
$$P(X = 1 \text{ or } X = 2) = 3/4$$

Cumulative distribution function: $F(x)$

$$F(x) = P(X \leq x), \text{ for all } x$$

$$F(1) = p(0) + P(1) = 3/4$$

$$F(1.5) = 1/4 + 1/2 = 3/4$$

$$F(2) = 1$$

$F(x)$ is a step function

In general,

$$F(x) = \sum_{t \leq x} p(t)$$

$$F(x) = \sum_{t} I(t \leq x) p(t)$$

Properties:

◦ *Monotone* : if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$

◦ $P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$

◦ $\lim_{x \to +\infty} F(x) = 1$

◦ $\lim_{x \to -\infty} F(x) = 0$

1. Bernouilli distribution: $B_r(.)$

$X \sim Bernouilli$, on $X = 0$ or $X = 1$

Prior probability of sucess is θ.

Bernouilli has: Two modes θ and $(1 - \theta)$ $(0 \leq \theta \leq 1)$

$$P(X = x|\theta) = B_r(x|\theta)$$

$$= \left\{ \begin{array}{c} \theta^x (1-\theta)^{1-x} \quad x = 0 \text{ or } 1 \\ 0 \text{ otherwise} \end{array} \right\}$$

◦ If θ is known: uncertainty about x

◦ If θ is not known: two uncertainties

Mean:

$$E(x|\theta) = \theta$$

Variance:

$$\begin{aligned} E(x^2|\theta) &= \theta \\ Var(x|\theta) &= E(x^2|\theta) - (E(x|\theta))^2 \\ &= \theta(1-\theta) \end{aligned}$$

2. Binomial: $B_i(.)$

$X \sim Binomial$ which means that x is defined on $0, 1, \ldots, n$

$$P(X = x|\theta, n) = B_i(x|\theta, n)$$

$$= \left\{ \begin{array}{cc} \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x} & x = 0, 1 \ldots, n \\ 0 \text{ otherwise} & \end{array} \right\}$$

$0 \leq \theta \leq 1$ and $n \geq 1$

In fact:

◦ If we have n iid Bernouilli trials $Y_1, \ldots, Y_n$,
◦ Each with ”success” probability θ

$$Y_i = \begin{array}{ll} 1 & \text{with probability } \theta \\ 0 & \text{with probability } 1-\theta \end{array}$$

then
$X = \sum_{i=1}^{n} Y_i \sim B_i(\theta, n)$

Mean:

$$E(X|\theta, n) = n\theta$$

$$Var(X|\theta, n) = n\theta(1-\theta)$$

Example 1:

X: is the number of females observed in a randomly chosen birth $n = 3$.
X: ranges from 0 to 3
θ : probability of a female birth, $\theta = 0.49$

$$P(X = 0|\theta, n) = 0.1327$$

$$P(X = 1|\theta, n) = ?$$

$$P(X = 2|\theta, n) = ?$$

$$P(X = 0 \text{ or } X = 2|\theta, n) = ?$$

HomeWork1:
simulate a series of random number from a binomial distribution using Winbugs or R or Splus or Matlab?

Approximation of the Binomial: Poisson

$$\left.\begin{array}{l}\theta \text{ is small}\\ n \text{ is large}\end{array}\right. \Rightarrow \textit{Binomial} \text{ becomes } \textit{Poisson} \text{ with rate } \lambda$$

In fact:

$\lambda = n\theta$

$$B_i(x = k|\theta, n) = \begin{pmatrix} k \\ n \end{pmatrix} \theta^k (1-\theta)^{n-k}$$

$$= n \times \frac{(n-1)\ldots(n-k+1)}{k!} \theta^k (1-\theta)^{n-k}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \ldots \times \frac{n-k+1}{n} \times \frac{\lambda^k}{k!}$$

$$\times (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-k}$$

$$\lim_{n\to\infty} B_i(k|\theta, n) = \frac{\lambda^k}{k!} \lim_{n\to\infty} (1 - \frac{\lambda}{n})^n$$

If we use the identity $(a+b)^n = \sum_{x=0}^{n} \begin{pmatrix} n \\ x \end{pmatrix} b^x a^{n-x}$ to show that

$$\begin{aligned} \lim_{n \to \infty} B_i(k|\theta, n) &= \frac{\lambda^k}{k!} \sum_{x=0}^{\infty} - \frac{(\lambda)}{x!}^{x} \\ &\simeq \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

3. Poisson: $Pois(.)$

$X \sim Pois(.|\lambda)$

Parameter: λ

$$Pois(X = x|\lambda) = \left\{ \begin{array}{cc} \frac{\lambda^x}{x!}e^{-\lambda} & \text{for } x = 0, 1, \ldots, \infty \\ 0 & \text{otherwise} \end{array} \right\}$$

Example 2:

We know that there is a 20% chance that a mouse exposed to UV light will develop skin cancer. We exposed 60 mice to UV.
What is the probability that 2 or more mice develop skin cancer?

x: number of mice that develop cancer
$\theta = 0.02$
$n = 60$
Here we use $B_i(x|\theta = 0.02, n = 60) = \begin{pmatrix} x \\ n \end{pmatrix} \theta^x (1-\theta)^{n-x}$.
In fact:

$$P(x \geq 2|\theta) = 1 - P(x \leq 1|\theta)$$
$$1 - P(x = 0|\theta) - P(x = 1|\theta) = 0.3381$$

If we use the poisson density function, we have the same result (n is large and θ is small)
Hint: $\lambda = n\theta = 60 \times 0.02 = 1.2$
In fact:

Approximation of the Binomial: Normal

$X \sim B_i(\theta, n)$

$X = \sum_{i=1}^{n} U_i$ (U_i are iid), $U_i \sim B_r(\theta)$

$E(U_i) = \theta, \qquad var(U_i) = \theta(1-\theta)$

$E(X) = n\theta, \qquad var(X) = n\theta(1-\theta)$

$$Z = \frac{X - n\theta}{[n\theta(1-\theta)]^{1/2}}$$

$$E(Z) = 0,\ Var(Z) = 1 \text{ (CLT)}$$

$$Z_{n\to\infty} \sim N(0,1)$$

If $X \sim B_i(\theta, n)$, given fixed x_1 and x_2

$$\begin{aligned} P(x_1 < X \leq x_2) &= \\ P\left(\frac{x_1 - n\theta}{[n\theta(1-\theta)]^{1/2}} < \frac{X - n\theta}{[n\theta(1-\theta)]^{1/2}} \leq \frac{x_2 - n\theta}{[n\theta(1-\theta)]^{1/2}}\right) \\ &= P(z_1 < Z \leq z_2) \\ &= P(Z \leq z_2) - P(Z \leq z_1) \\ &\simeq \Phi(z_2) - \Phi(z_1) \end{aligned}$$

where Φ is the c.d.f of the standard normal.

If we use this approximation for the mouse example

$$
\begin{aligned}
&P(X \geq 2) \\
&\qquad = P\left(\frac{X - n\theta}{[n\theta(1-\theta)]^{1/2}} \geq \frac{2 - n\theta}{[n\theta(1-\theta)]^{1/2}}\right) \\
&\qquad = P(Z \geq 0.7377) \\
&\qquad = 1 - P(Z < 0.7377) \\
&\qquad \simeq 0.230 \text{ very different from } 0.33
\end{aligned}
$$

So the normal approximation is not adequate here.

Condition: $\min(n\theta, n(1-\theta)) \geq 5)$
Here $n\theta = 1.2$

Simulating a discrete random variable:

Suppose X has a discrete distribution

$P(X = x_i) = p_i \quad i = 1,..,4$

$\sum_{i=1}^{4} p_i = 1$

1. $U \sim U_n(0,1)$

$$X = \left\{ \begin{array}{cc} x_1 & \text{if } U \leq p_1 \\ x_2 & \text{if } p_1 \leq U \leq p_1 + p_2 \\ x_3 & \text{if } p_1 + p_2 \leq U \leq p_1 + p_2 + p_3 \\ x_4 & \text{if } p_1 + p_2 + p_3 \leq U \leq 1 \end{array} \right\}$$

Test that $P(X = x) = p_3$
In fact:

$$\begin{aligned} P(X = x_3) \\ &= P(p_1 + p_2 \leq U \leq p_1 + p_2 + p_3) \\ &= P(U \leq p_1 + p_2 + p_3) - P(U \leq p_1 + p_2) \\ &= \int_0^{p_1+p_2+p_3} du - \int_0^{p_1+p_2} du \\ &= p_3 \end{aligned}$$

This is called ”Inverse transform method”

1.2 Univariate continuous distribution

X: random variable taking infinite values in R
X is continuous and its c.d.f $F(X)$ is continuous.
$P(X = x) = 0$

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} p(t)dt \text{ for } -\infty < x < +\infty \end{aligned}$$

◦ If $F(x)$ is differentiable, then $\frac{d}{dx}F(x) = p(x)$
◦ Two random variables may have the same c.d.f or p.d.f
Example: ????

Properties of p.d.f:

◦ $p(x) \geq 0$ for all x

◦ $\int_{-\infty}^{+\infty} p(x)dx = 1$

◦ $x \in (a,b)$, $P(a < x \leq b) = \int_a^b P(x)dx = F(b) - F(a)$

Homework 2: prove that

$$P(a < x < b) = P(a < x \leq b) = P(a \leq x \leq b) = P(a < x \leq b)$$

◦ $P(X = x) = \lim_{\epsilon \to 0} \int_{a-\epsilon}^{a+\epsilon} p(x)dx = 0$

◦$P(x \leq X \leq x + \Delta) = \int_x^{x+\Delta} p(x)dx \cong \Delta p(x)$

◦$If\ A \subset R$, then

$$\begin{aligned} p(X \in A) &= \int_A p(x)dx \\ &= \int I(x \in A)p(x)dx = E[I(x \in A)] \end{aligned}$$

◦Expectation of a function g with respect to a probability in R is

$$\int g(x)p(x)dx$$

1. Kernel:

A function $f(x)$ such that

$$f(x) \propto p(x)$$

is called the *kernel* of $p(x)$

if $p(x) \propto f(x)$ then

$$p(x) = C \times f(x) = \frac{f(x)}{\int_{-\infty}^{+\infty} f(x)dx}$$

Example:

$$X \sim N(\mu, \sigma^2)$$

The kernel is

$$f(x) = \exp\left(\frac{-(x-\mu)^2}{2\sigma^2} \right)$$

then

$$C = \frac{1}{\int_{-\infty}^{+\infty} f(x)dx} = (2\pi\sigma^2)^{-1/2}$$

Homework 3:

Prove the above result?

2. Uniform

$$X \sim U_n(a,b)$$

$$\begin{aligned} p(x|a,b) &= U_n(x|a,b) \\ &= \left(\begin{array}{cc} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{array} \right) \end{aligned}$$

$$F(x|a,b) = \left(\begin{array}{cc} 0 & x < a \\ \int_a^x U_n(x|a,b)dx = \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{array} \right)$$

Mean and Variance:

$$E(X|a,b) = \frac{a+b}{2}$$

$$\begin{aligned} Var(X|a,b) &= E(X^2|a,b) - E(X|a,b)^2 \\ &= \frac{(a+b)^2 - ab}{3} - (\frac{a+b}{2})^2 \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Homework 4:

Write a program in your preferred language for simulating a random variable having a uniform distribution

Simulation of a continuous variable:

We want to generate a continuous random variable X

X has a c.d.f function $F(x)$

1. Generate $U \sim U_n\ (0,1)$
2. $X = F_X^{-1}(U)$

Test that:

$$P(X \leq x) = F_X(x)$$

In fact:

$$\begin{aligned} P(F_X^{-1}(U) &\leq x) \\ &= P(U \leq F_X(x)) \\ &= \int_0^{F_X(x)} du \\ &= F_X(x) \end{aligned}$$

Homework 5: Write a program that simulate a random variable from the power exponential defined as the following:

$$f(x; \mu, \mathbf{\Sigma}, \beta) = \frac{p\Gamma(\frac{p}{2})}{\pi^{p/2}\Gamma(1+\frac{p}{2\beta})2^{1+\frac{p}{2\beta}}} |\mathbf{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\left[(x-\mu)' \, \mathbf{\Sigma}^{-1}(x-\mu)\right]^{\beta}\right\},$$

3. Beta

$X \sim \ Beta \Rightarrow X \in (0,1)$

$$B_e(x|a,b) = \begin{pmatrix} C\,x^{a-1}(1-x)^{b-1} & x \in [0,1] \\ 0 & otherwise \end{pmatrix}$$

$$a > 0,\ b > 0$$

$$C = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$\begin{aligned} \Gamma(\alpha) &= \int_0^{+\infty} x^{\alpha-1}e^{-x}dx,\ \alpha > 0 \\ &= (\alpha-1)! \end{aligned}$$

Mean:

$$\begin{aligned} E(X|a,b) &= C\int_0^1 x[x^{a-1}(1-x)^{b-1}]dx \\ &= C\int x^{(a+1)-1}(1-x)^{b-1}dx \\ &= C\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\ &= \frac{a}{a+b} \end{aligned}$$

Variance:

$$\begin{aligned} Var(X|a,b) &= E(X^2|a,b) - E^2(X|a,b) \\ &= \frac{ab}{(a+b)^2(a+b+1)} \end{aligned}$$

4. Gamma

$X \sim Gamma$

$$\begin{aligned} P(x|a,b) &= G_a(x|a,b) \\ &= C\,x^{a-1}e^{-bx} \end{aligned}$$

$$\left(\begin{array}{cc} C = \frac{b^a}{\Gamma(a)} & \begin{array}{c} a : shape \\ b : scale \end{array} \end{array} \right)$$

- ◦ When $a = 1, \quad p(x|1,b) = b\exp(-bx),\ x > 0,$ then this is an exponential distribution
- ◦ When $a = \frac{v}{2},\ b = \frac{1}{2},$ then this is a central Chi-square distribution

$$p(x|v) = \left(\begin{array}{cc} Cx^{v/2-1}\exp(-x/2) & v > 0 \\ C = \frac{(1/2)^{v/2}}{\Gamma(v/2)} & x > 0 \end{array} \right)$$

$E(X^k|a,b) = \int_0^{+\infty} x^k \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)dx$

$= \frac{b^a}{\Gamma(a)} \int_0^{+\infty} x^{k+a-1} \exp(-bx)dx$

If $X = Y^2$, $dx = 2ydy$

$E(X^k|a,b) = \frac{2b^a}{\Gamma(a)} \int_0^{+\infty} y^{2(k+a)-1} \exp(-by^2)dy$

$= \frac{b^a}{\Gamma(a)} b^{-2(k+a)/2} \Gamma(\frac{2(k+a)}{2})$

Mean:

$E(X|a,b) = \frac{\Gamma(a+1)}{b\Gamma(a)} = \frac{a}{b}$

$E(X^2|a,b) = \frac{\Gamma(a+2)}{b^2\Gamma(a)} = \frac{a(a+1)}{b^2}$

$Var(X|a,b) = \frac{a}{b^2}$

5. Normal

$X \sim Normal$

$$P(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$for \ -\infty < x < +\infty$$

We will show that:

1. $E(X) = M'(0)$
2. $E(X^2) = M''(0)$
3. ....$E(X^k) = M^{(k)}(0)$

Moment generating function is:

$$M(t) = E[\exp(tX)]$$

In fact

$M(t) = E(\exp(tX)) \ = 1 + tE(X) + t^2\frac{E(X^2)}{2!} + \ldots + t^k\frac{E(X^k)}{k!}$

$\frac{\partial E(\exp(tX))}{\partial t} = E(X) + 2t\frac{E(X^2)}{2!} + \ldots + kt^{(k-1)}\frac{E(X^k)}{k!}$

$\frac{\partial E(\exp(tX))}{\partial t}_{t=0} = E(X)$

Finding mean and variance:

$$\begin{aligned} M(t) &= \int_{-\infty}^{+\infty} \exp(tx)p(x)dx \\ &= \int_{-\infty}^{+\infty} \exp(tx)\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)dx \\ &= \frac{\exp(\mu t + \frac{\sigma^2 t^2}{2})}{\sigma\sqrt{2\pi}} \end{aligned}$$

$$\begin{aligned} \frac{\partial E(\exp(tX))}{t} &= \frac{\mu + \sigma^2 t}{\sigma\sqrt{2\pi}} \exp(\mu t + \frac{\sigma^2 t^2}{2})(0) \\ &= \mu \end{aligned}$$

$E(X^2) = \mu^2 + \sigma^2$

6. Student

$$p(x|\mu,\sigma^2,\nu)$$

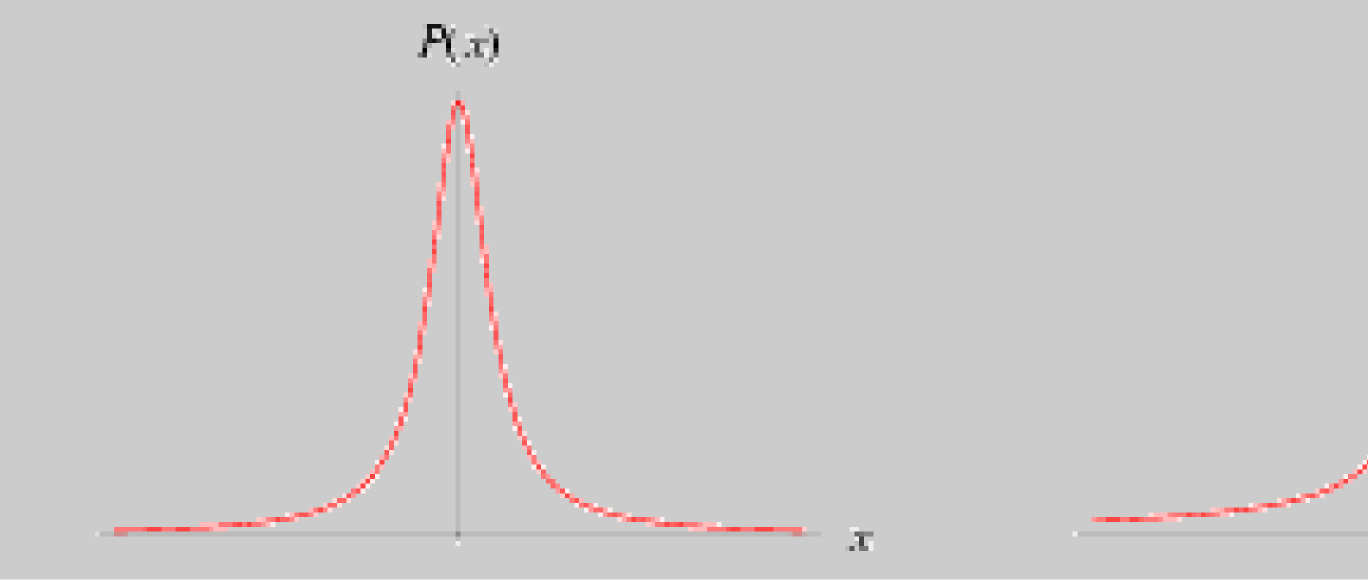
$$= \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi\sigma}}\left[1+\frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-(\nu+1)/2}$$

This has three parameters:
μ : its mean
ν : its degrees of freedom
σ : its scale .

The variance of a t distribution $Var(X|\mu,\sigma^2,\nu) = \frac{\nu}{(\nu-2)}\sigma^2$

- if $\nu \leq 2;\ var \rightarrow \infty$
- if $\nu \rightarrow \infty;$ process is normal
- if $\nu = 1;$ Cauchy distribution (mean and higher moment do not exist)

$$f(x) = \frac{1}{\pi}\frac{b}{(x-m)^2+b^2}$$



We will later that (given μ and s_i^2)

$$(X_i|\mu,s_i^2,\nu)\sim N(\mu,s_i^2)$$

and
if $(s_i^2|\nu,\sigma^2) \sim \nu\sigma^2\chi_\nu^{-2}$, a scaled inverted chi-square with parameters ν and σ^2

$$p(x_i|\mu, \sigma^2, \nu) = \int p(x_i|\mu, s_i^2) p(s_i^2|\nu, \sigma^2) ds_i^2$$

is a t-distribution. It is a mixture of normals with the same mean μ and different variance s_i^2.

Simulation of a t-random:

1. simulate $s_i^2 \sim \nu\sigma^2\chi_\nu^{-2}$
2. simulate $x_i \sim N(\mu, s_i^2)$

Homework 6:

Use Winbugs, or R, or Splus or Matlab to simulate from a t distribution using:

1. Straightfoward method
2. method above
3. Compare your result

1.3: Multivariate probability distribution

Independency:
$X_1, X_2, \ldots., X_n$ are iid if:

$$P(x_1 \ldots x_n) = \prod_{i=1}^{n} p(x_i)$$

$p(x_i)$: marginal probability

$$p(x_i) = \int p(x_1 \ldots x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$$

Exchangeability:
if

$$p(x_1 \ldots x_n) = p(x_{p(1)} \ldots . x_{p(n)})$$

Bivariate case:

$$F(x,y) = p(X \le x; Y \le y) \text{ continuous}$$
$$F(x,y) = p(X = x; Y = y) \text{ discrete}$$

p.d.f:

$$p(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y},\ (x,y) \in R^2$$

$$\begin{aligned} p((X,Y) \in A) &= \int\int_A p(x,y)dxdy \\ &= \int\int I[(x,y) \in A]p(x,y)dxdy \end{aligned}$$

Conditional:

- ◦ $p(X|Y=y) = p(x|y) = \frac{p(x,y)}{p(y)}, p(y) > 0$
- ◦ if $p(y) = 0 \Rightarrow p(x|y) = p(x)$
- ◦ $\int p(x|y)dx = \int \frac{p(x,y)dx}{p(y)} = 1$
- ◦ $p(x|y) \propto p(x,y)$

◦ $p(x,y) = \begin{pmatrix} p(x)p(y) \\ p(x|y) = p(x) \end{pmatrix}$

if x and y are independent

If $X_1, \ldots X_n$ are iid

$$p(X_1 \ldots X_k | rest) = p(X_1 \ldots X_k);\ k < n$$

Marginals:

Discrete:

$$p(X = x) = \sum_y p(x,y) = \sum_y p(x|y)p(y)$$

Continuous:

$$p(x) = \int p(x,y)dx = \int p(x|y)p(y)dy$$

$$\begin{aligned} p(x,y) &= c \times f(x,y) \\ &\Rightarrow \\ p(x,y) &= \frac{f(x,y)}{\int\int f(x,y)dxdy} \\ &\Rightarrow \\ c^{-1} &= \int\int f(x,y)dxdy \end{aligned}$$

Example 1: Aitken's integral

$$(X, Y) \sim N(\mu, V)$$

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, V = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$$

$p(X, Y|\mu, V) =$

$$\frac{1}{\sqrt{2\pi}\,|V|^{1/2}} \exp \frac{-1}{2} (X - \mu_X, Y - \mu_Y)^t V \begin{pmatrix} X - \mu_X \\ Y - \mu_Y \end{pmatrix}$$

kernel of the density is $f(x,y)$

$$\exp\left\{ -\frac{1}{2} (X - \mu_X, Y - \mu_Y)^t V \begin{pmatrix} X - \mu_X \\ Y - \mu_Y \end{pmatrix} \right\}$$

$$\begin{aligned} c^{-1} &= \int\int f(x,y) dxdy \\ &= \int\int \exp\left\{ -\frac{1}{2} (X - \mu_X, Y - \mu_Y)^t V \begin{pmatrix} X - \mu_X \\ Y - \mu_Y \end{pmatrix} \right\} dxdy \\ &= 2\pi |V|^{1/2} \end{aligned}$$

Example 2:

$$p(x,y) = \left(\begin{array}{cc} 6xy^2 & 0 < x < 1; 0 < y < 1 \\ 0 & otherwise \end{array} \right)$$

we have:

$$\int\int p(x,y)dxdy = 1$$

$$p(y) = 3y^2$$

$$p(x) = 2x$$

∘$p(y|x) = \frac{p(x,y)}{p(x)} = 3y^2; 0 < y < 1$

∘$p(x|y) = 2x$

So X and Y are independent

∘$E(Y|X = x) = \int yp(y|x)dy = 3/4$

∘$Var(Y|X = x) = E(Y^2|x) - E^2(Y|x) = 3/80$

Multinomial Distribution:

◦ Suppose $C_1, \ldots, C_k$ are k mutually exclusive and exhaustive classes or outcomes.
◦ n experiments (independent trials)
◦ One of the k outcomes is possible
◦ P(outcomes)$=p_i$

$$\begin{aligned} p(X_1 &= x_1, \ldots X_k = x_k | x_1 + \ldots + x_k = n) = \\ &= \frac{n!}{x_1! x_2! \ldots . x_k!} p_1^{x_1} \ldots .. p_k^{x_k} \\ &= \frac{n! p_1^{x_1} \ldots p_{k-1}^{x_{k-1}} (1 - p_1 \ldots - p_{k-1})^{(n - x_1 \ldots . - x_{k-1})}}{x_1! \ldots x_{k-1}! (n - x_1 - \ldots - x_{k-1})!} \end{aligned}$$

Mean and variance:

$$E(X_i) = np_i$$
$$Var(X_i) = np_i(1 - p_i)$$
$$Cov(X_i, X_j) = -np_ip_j;\ i \neq j$$

if X_i: count of outcome (i) and X_{-i} : "non-i"
Then, the marginal

$$X_i | X_{-i} \sim B_i(X_i | p_i, n)$$

$$X_{-i} |\ X_i \sim M_u(q_1, \ldots .q_{i-1}, q_{i+1}, \ldots q_k, n - X_i)$$

where $q_j = p_j/(1 - p_i)$

$$X_{-i,-j} | X_i, X_j \sim$$
$$M_u(r_1, \ldots, r_{i-1}, r_{i+1}, \ldots r_{j-1}, r_{j+1}, \ldots, n - X_i - X_j)$$

where $r_s = p_s/(1 - p_i - p_j)$

Normal approximation:

$(n \longrightarrow \infty)$

$$E(X_i) = np_i,\ i = 1, \ldots, k$$

$V = Var(X) =$

$$n \times \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \ldots & -p_1p_k \\ -p_2p_1 & & & \\ \vdots & & \ddots & \\ -p_kp_1 & & & p_k(1-p_k) \end{pmatrix}$$

where $npi(1 - p_1 - p_2 \ldots - p_k) = 0$

Stuart and Ord (1991) Book: Kendall’s advanced theory of statistics

$Rank(V) = k - 1,$ it means that a generalized inverse of V is

$$V^{-1} = \begin{pmatrix} \tilde{V}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\tilde{V} = \begin{pmatrix} \frac{1}{p_1}+\frac{1}{p_k} & \frac{1}{p_2} & \cdots & \frac{1}{p_k} \\ & \frac{1}{p_2}+\frac{1}{p_k} & & \\ \vdots & & \ddots & \\ \frac{1}{p_k} & & & \frac{1}{p_{k-1}}+\frac{1}{p_k} \end{pmatrix}$$

We have

$$\exp[-(X-\mu)^t V^{-1}(X-\mu)]$$
$$= \exp[(X-np)^t V^{-1}(X-np)]$$

where

$$p^t = (p_1, \ldots, p_k)$$

when $n \longrightarrow +\infty$

$$(X-np)^t V^{-1}(X-np)$$
$$\sim\chi_\nu^2(Central)$$

where $\nu = tr(V^{-1}V) = k-1$

Simulating from a Multinomial:

$x = (x_1, \ldots, x_d)$ with a $p.d.f\ \ f(x)$

$$f(x) = f(x_1)f(x_2|x_1).\ldots.f(x_d|x_1,\ldots x_{d-1})$$

If the marginal and the conditional are known, this method will allow to reduce a multivariate to a $d - univariate$ problem.

Example: $(X_1, X_2, X_3, X_4) \sim M_u(p_1, p_2, p_3, p_4, n)$

1. $X_1 \sim B_i(p_1, n)$
2. $X_2|X_1 \sim B_i(\frac{p_2}{1-p_1}, n - x_1)$
3. $X_3|X_1, X_2 \sim B_i(\frac{p_3}{1-p_1-p_2}, n - x_1 - x_2)$

$Set\ x_4 = n - x_1 - x_2 - x_3$

$If\ n = 0;\ B_i(p, 0) = 0$

Homework 6: Use Winbugs, R or Matlab to simulate from a 4-multinomial using the method above

Dirichlet distribution:

is a generalization of *Beta*

$X = (X_1,..,X_k), \qquad X_1 \geq 0,....,X_k \geq 0$
$\sum_{i=1}^{k} X_i = 1$

$$(X|\alpha_1,..,\alpha_k) \sim D_i(\alpha_1,...,\alpha_k)$$

$\alpha_j > 0$

$$p(x|\alpha_1,..,\alpha_k) = \frac{\Gamma(\alpha_1 +...\alpha_k)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_k)} \prod_{i=1}^{k} x_i^{\alpha_i - 1}$$

Simulation:

1. Simulate k

$$y_i \sim G_a(y_i|\alpha_i, 1)$$

2.

$$x_i = \frac{y_i}{\sum_{j=1}^{k} y_j}, \ j = 1, \ldots, k$$

Use winbugs to simulate:

1. directly from a dirichlet
2. using the method above

d-dimensional uniform distribution

$$X \sim U_n(0,1)^d$$

if

$$p(x) = \begin{matrix} 1 & \text{if } x \in [0,1]^k \\ 0 & \text{otherwise} \end{matrix}$$

Multivariate Normal distribution:

$$\begin{aligned} & p(y|m, V) \\ &= |2\pi V|^{-1/2} \exp\left[-\frac{1}{2}(y-m)^t V^{-1}(y-m)\right] \end{aligned}$$

$$m = E(y|m, V) = \int y p(y|m, V) dy$$

$$\begin{aligned} V &= E[(y-m)(y-m)^t] \\ &= \int (y-m)(y-m)^t p(y|m, V) dy \\ &= \int y y^t p(y|m, V) dy - m m^t \end{aligned}$$

if

$$y = [y_1^t, y_2^t]^t$$

a decomposition of $y = (y_1, \dots, y_n)$
and

$$m = [m_1^t, m_2^t]^t$$

and

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

then

$$\begin{aligned} p(y_1|m_1, V_{11}) &= \int_{-\infty}^{+\infty} p(y_1, y_2|m_1, V_{11}) dy_2 \\ &= (2\pi)^{-n_1/2}|V_{11}|^{1/2} \\ &\quad \exp\left[-\frac{1}{2}(y_1 - m_1^t)^t V_{11}^{-1}(y_1 - m_1)\right] \end{aligned}$$

$$(y_1|y_2) \sim N[E(y_1|y_2), Var(y_1|y_2)]$$

$$\begin{aligned} &p(y_1|m_1, V_{11}) \\ &= (2\pi)^{-n_1/2}|V(y_1|y_2)|^{1/2} \\ &\quad \times \exp\left[-\frac{1}{2}(y_1 - E(y|y))^t V(y_1|y_2)(y_1 - E(y_1|y_2))\right] \end{aligned}$$

$$\begin{aligned} E(y_1|y_2) &= m_1 + V_{12}V_{22}^{-1}(y_2 - m_2) \\ Var(y_1|y_2) &= V_{11} - V_{12}V_{22}^{-1}V_{21} \end{aligned}$$

Y_1 and Y_2 are independent if $V = \sigma^2 I.$

$$(y - m)^t V^{-1} (y - m) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - m)^2$$

$$|2\pi V|^{-1/2} = |2\pi\sigma^2 I| = (2\pi\sigma^2)^{-n/2}$$

$$p(y|m, V) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{\sum_{i=1}^{n} (y_i - m)^2}{2\sigma^2} \right]$$

$$= \prod_{i=1}^{n} p(y_i|m, \sigma^2)$$

When $rank(V) = k < n$ (singularity) then

$$p(y|m, V) = \frac{(2\pi)^{-k/2}}{(\lambda_1 \ldots \lambda_k)^{1/2}}$$

$$\exp\left[-\frac{1}{2} (y - m)^t V^{-1} (y - m) \right]$$

$V^{-1}$is generalized inverse of V

$\lambda_1, \ldots, \lambda_k$ are non-zero eigenvalues of V

see Searle 1971; Rao(1973); Anderson (1984); Mardia(1979)

Moment Generating function

$$y \sim N(m, V)$$

$$M(t) = E(\exp(t^t y))$$

$$\begin{aligned} M(t) &= \frac{1}{|2\pi V|^{1/2}} \times \\ \int & \quad \exp\left[-\frac{1}{2}(y-m)^t V^{-1}(y-m)\right] dy \\ &= \exp\left(t^t m + \frac{t^t V t}{2}\right) \end{aligned}$$

$$E(y) = \frac{\partial M(t)}{\partial t} = (m + Vt)\exp(t^t m + \frac{t^t V t}{2})$$

$t = 0 \Rightarrow E(y) = m$

$$\begin{aligned} \frac{\partial^2 M(t)}{\partial t \partial t^t} &= \exp(t^t m + \frac{t^t V t}{2}) \\ &\quad \times [V + (m + Vt)(m + Vt)^t] \end{aligned}$$

if $t = 0 \Rightarrow E(yy^t) = V + mm^t$

$\Rightarrow E(yy^t) - E(y)E(y^t) = V$

Linear function of normally distributed random:

A linear transformation of normal is normal
let

$$
\begin{aligned}
x &= \alpha + \beta^t y \\
&\quad y \sim Normal \\
&\quad \alpha, \text{ and } \beta \text{ are known}
\end{aligned}
$$

Moment generator function:

$$
\begin{aligned}
& E[\exp(\alpha + \beta^t y)t] \\
&= \exp(\alpha t)E(\exp(t\beta^t y)) \\
&= \exp(\alpha t)M(t^*) \\
&= \exp\left[\, t(\alpha + \beta^t m) + \frac{t^2(\beta^t V\beta)}{2} \,\right]
\end{aligned}
$$

where $t^* = t\beta$
which is the moment generator of a Normal density function with

$$
\begin{aligned}
mean &= (\alpha + \beta^t m) \\
Var &= \beta^t V\beta
\end{aligned}
$$

Central limit theorem

Example: Infinitesimal model:

Def: additive value of a quantitative trait is the result of the sum of value at each of infinite number of trait.

We will show that the use of CLT means that the additive value $Y = \sum_{i=1}^{n} Y_i$ follows a normal distribution independently of the individual distribution

$Y = \sum_{i=1}^{n} Y_i \qquad E(Y) = \sum_{i=1}^{n} E(Y_i) = \sum \mu_i = \mu$

$Var(Y) = \sum \sigma_i^2 = \sigma^2$

$$Z = \frac{Y - \mu}{\sigma}$$

$$\begin{aligned} M_Z(t) &= E(\exp(tZ)) \\ &= E(\exp \sum_{i=1}^{n} tZ_i) \end{aligned}$$

where $Z_i = \frac{Y_i - \mu_i}{\sigma}$ and $\sum_i Zi = Z$

$$\begin{aligned} M_Z(t) &= \int \exp(\sum_i tZ_i) p(z_1, .., z_n) dz \\ &= \prod E\left\{\exp \frac{t(Y_i - \mu_i)}{\sigma}\right\} \\ & \prod_i M_i(\frac{t}{\sigma}) \end{aligned}$$

where $M_i(t)$ is the moment generator of $(y_i - \mu_i)$

$\Rightarrow$

$$\begin{aligned} M_i(\frac{t}{\sigma}) &= E\left(\exp \frac{t(Y_i - \mu_i)}{\sigma}\right) \\ &= 1 + t \underbrace{\frac{E(Y_i - \mu_i)}{\sigma}}_{=0} + t^2 \frac{E(Y_i - \mu_i)^2}{\sigma^2} + \underbrace{\ldots}_{=0} \\ &\approx 1 + \frac{t\sigma_i^2}{2\sigma^2} + \ldots.. \\ \log(M_Z(t)) &= \sum_{i=1}^{n} \log\left(1 + \frac{t^2\sigma_i^2}{2\sigma^2}\right) \end{aligned}$$

which means that

(using expansion of $log(1 + x)$ at $x = 0$ (near 0, we have $log(1 + x) \approx 0)$

$$M_Z(t) \approx \exp \sum_{i=1}^{n} \frac{t^2\sigma_i^2}{2\sigma^2} = \exp(\frac{t^2}{2})$$

this is the moment of a Normal

$Z \sim N(0,1)$ so $Y = \mu + Z\sigma^2 \sim N(\mu,\sigma^2)$

Quadratic forms on Normal variable: Chi-square:

$$Y \sim N(m, V)$$

$y^t Q\, y$ is a quadratic form

If QV is idempotent, then

$$y^t Q\, y \sim \chi^2\left(\underbrace{rank(Q);}_{=df} \qquad \underbrace{\frac{1}{2} m^t Q m}_{\text{non centrality parameter}}\right)$$

$$E(y^t Q y) = m^t Q m + tr(QV)$$

$$Var(y^t Q y) = 4 m^t QVQm + 2tr(QV)^2$$

Example:

if

$$y \sim N(X\beta, \sigma_e^2 I)$$

then MLE(σ_e^2) is

$$\begin{aligned}\hat{\sigma}_e^2 &= \frac{1}{n}(y - X\widehat{\beta})^t (Y - X\widehat{\beta}) \\ &= \frac{1}{n} y^t Q y\end{aligned}$$

where:

- $\hat{\beta} = (X^t X)^{-1} X^t y$
- $Q = [I - X(X^t X)^{-1} X^t],\ Q$ is idempotent.

We have

$$\begin{aligned}\hat{\sigma}_e^2 &= \frac{\sigma_e^2}{n} \frac{y^t Q y}{\sigma_e^2} \\ &= \frac{\sigma_e^2}{n} y^t Q \left(\frac{I}{\sigma_e^2}\right) Q y \\ &= \frac{\sigma_e^2}{n} y^{*t} \left(\frac{I}{\sigma_e^2}\right) y^*\end{aligned}$$

where $y^* = Qy$

so

$$\hat{\sigma}_e^2 \propto y^{*t}(I\sigma_e^{-2}) y^*$$

$$y^* \sim N\,(0, Q\sigma_e^2)$$

$$y^{*t}\left(\frac{I}{\sigma_e^2}\right) y^* \sim \chi^2_{df}(\text{central})$$

$$df = rank(Q) = tr(Q) = tr[I - X(X^t X)^{-1} X^t] = n - tr[X(X^t X)^{-1} X^t] - n - p$$

and

$$\hat{\sigma}_e^2 \sim \frac{\sigma_e^2}{n} \chi_{n-p}^2$$

So the MLE of σ_e^2 has a scaled chi-square distribution.

Wishart and Inverse Wishart:

Wishart:

$Y = (Y_1, \ldots, Y_n)$

$$Y \sim N(0, \Sigma)$$

$M = Y^t Y$ of dimension p by p

$$M \sim W_p(\Sigma, n)$$

p.d.f

$$\begin{aligned} p(M|\Sigma, n) &= \frac{|M|^{(n-p-1)/2} \exp[-\frac{1}{2} tr(\Sigma^{-1} M)]}{2^{np/2} \pi^{p(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+1-i)]} \\ &\propto |M|^{(n-p-1)/2} \exp\left[\frac{1}{2} tr(\Sigma^{-1} M)\right] \end{aligned}$$

$E(M|\Sigma, n) = n\Sigma$

Inverse Wishart:

$T = M^{-1}$

$$T \sim W_p^{-1}(\Sigma, n)$$

$$p(T|\Sigma, n) = \frac{|M|^{(n+p+1)/2} \exp[-\frac{1}{2} tr(\Sigma^{-1} T^{-1})]}{2^{np/2} \pi^{p(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+1-i)]}$$

$$\propto |T|^{(n+p+1)/2} \exp\left[\frac{1}{2} tr(\Sigma^{-1} M^{-1})\right]$$

$$E(T|\Sigma, n) = \frac{\Sigma^{-1}}{(n-p-1)}$$

Particular case:

if $p = 1, T = \sigma^2$ and $\Sigma^{-1} = s$

$$p(\sigma^2|s, n) \propto (\sigma^2)^{-(\frac{n}{2}+1)} \exp(-\frac{s}{2\sigma^2})$$

$$\sim \chi^2(s, n)$$

$$\sim IG(n/2, s/2)$$

if $s = ns^*$;

- s is called prior sum of squares
- s^* is a value of σ^2 viewed as likely when n is large.
- n is the degree of belief in "s^*"

Properties of wishart and inverse wishart:

Let

$$M_{(2,2)} \sim W_2(V, n),$$

V is scale and

$$T = M^{-1} \sim W_2^{-1}(V, n)$$

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}; V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$$

Then:

$$\begin{aligned} M_{11} &= (T_{11} - T_{12}^2 T_{22}^{-1})^{-1} \\ M_{12} &= -T_{12}/(T_{11}T_{22} - T_{12}^2) \\ M_{22} &= (T_{22} - T_{12}^2 T_{11}^{-1})^{-1} \end{aligned}$$

We define:

$$\begin{aligned} X_1 &= M_{11} \\ X_2 &= M_{11}^{-1} M_{12} \\ X_3 &= M_{22} - M_{12}^2 M_{11}^{-1} \end{aligned}$$

then

$$\begin{aligned} T_{11} &= X_1^{-1} + X_2^2 X_3^{-1} \\ T_{12} &= -X_2 X_3^{-1} \\ T_{22} &= X_3^{-1} \end{aligned}$$

and

$$\begin{aligned} X_1 &\sim W_1(V_{11}, n) \\ X_2 | X_1 &\sim N(V_{11}^{-1} V_{12}, x_1^{-1} V_{22.1}) \\ X_3 &\sim W_1(V_{22.1}, n-1) \end{aligned}$$

$$V_{22.1} = V_{22} - V_{12}^2 V_{11}^{-1}$$

Simulating from $W_2(V,n)$:

1. Simulate X from a univariate Wishart, inverse Gamma or Normal
2. Compute T from X which has a $W_2^{-1}(V,n)$
3. Invert T to obtain a draw from $W_2(V,n)$

Simulation of a Wishart:

$$\Sigma \sim W_p(S,n)$$

1. Compute cholesky factorization

$$S = L^t L;\ L^t \text{ is lower triangular}$$

2. Construct $T = \begin{pmatrix} t_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ t_{p1} & \cdots & t_{pp} \end{pmatrix}$

$$\begin{aligned} t_{ii} &= \sqrt{\chi^2_{n+1-i}} \\ t_{ij} &\sim N(0,1), i > j \\ t_{ij} &= 0, i < j \end{aligned}$$

3. Compute $L^t T T^t L \sim W_p(S,n)$
4. $(L^t T T^t L)^{-1} \sim W_p^{-1}(S,n)$

How to choose the scale V in general:

If

$$C \sim W^{-1}(V, n)$$

Choose $V,$ such that

$$E(C|V, n) = \frac{V^{-1}}{n-p-1}$$

so we set

$$V^{-1} = (n-p-1)\,\tilde{E}(C|V, n)$$

where $\tilde{E}$ is some reasonable value chosen a priori.

Multivariate t-student:

suppose

$$y \sim N(\mu, \Sigma/\omega)$$

$$\omega \sim G_a(\upsilon/2, \upsilon/2)$$

$$\begin{aligned} p(y, \omega|\mu, \Sigma, \upsilon) &= p(y|\mu, \Sigma, \omega)p(\omega|\upsilon) \\ &= |2\pi\left(\frac{\Sigma}{\omega}\right)|^{-1/2} \\ &\quad \times \exp\left(-\frac{1}{2}(y-\mu)^t\left(\frac{\Sigma}{\omega}\right)^{-1}(y-\mu)\right) \\ &\quad \times \frac{\left(\frac{\upsilon}{2}\right)^{\upsilon/2}}{\Gamma\left(\frac{\upsilon}{2}\right)}\omega^{\frac{\upsilon}{2}-1}\exp\left[-\frac{\upsilon\omega}{2}\right] \end{aligned}$$

the marginal of y is

$$\begin{aligned} &\int p(y, \omega|\mu, \Sigma, \upsilon)d\omega \\ &= |2\pi\Sigma|^{-1/2}\frac{\left(\frac{\upsilon}{2}\right)^{\upsilon/2}}{\Gamma\left(\frac{\upsilon}{2}\right)} \\ &\quad \times \int \omega^{\frac{n+\upsilon}{2}-1}\exp\left(-\frac{\omega}{2}(y-\mu)^t\Sigma^{-1}(y-\mu)+\upsilon\right)d\omega \end{aligned}$$

the quantity inside the integrale is the kernel of a

$$G_a\left(\frac{n+\upsilon}{2}; \frac{1}{2}(y-\mu)^t\Sigma^{-1}(y-\mu)+\upsilon\right)$$

which means that it is:

$$\frac{\Gamma\left(\frac{n+\upsilon}{2}\right)}{\left[\frac{1}{2}(y-\mu)^t\Sigma^{-1}(y-\mu)+\upsilon\right]^{\frac{n+\upsilon}{2}}}$$

so

$$y|\mu, \Sigma, \upsilon \sim t(\mu, \Sigma, \upsilon)$$

Mean and variance:

$$\begin{aligned} E(y|\mu,\Sigma,\upsilon) &= E_{\omega}(y|\mu,\Sigma,\upsilon) \\ &= E_{\omega}(\mu) = \mu \end{aligned}$$

$$\begin{aligned} & Var(y|\mu,\Sigma,\upsilon) \\ &= E_{\omega}[Var(y|\mu,\Sigma,\upsilon)] + Var[E_{\omega}(y|\mu,\Sigma,\upsilon)] \\ &= E_{\omega}[Var(y|\mu,\Sigma,\upsilon)] \\ &= E_{\omega}\left(\frac{\Sigma}{\omega}\right) = \Sigma E_{\omega}\left(\frac{1}{\omega}\right) \end{aligned}$$

Marginal and conditional:

The marginal and conditional are also t-distributed
suppose that we partition the data such that

$$\left(\begin{matrix} y_1 \\ y_2 \end{matrix} \; |\mu,\Sigma,\omega)\right) \sim N\left(\begin{matrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \\ \frac{1}{\omega}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{matrix}\right)$$

Prove that:

$$\begin{aligned} & E(y_1|y_2,\mu,\Sigma,\upsilon) \\ &= \mu_1 + \Sigma_{12}(\Sigma_{22})^{-1}(y_2 - \mu_2) \end{aligned}$$

and that

$$\begin{aligned} & E(y_1|y_2,\mu,\Sigma,\upsilon) \\ &= \frac{\upsilon}{\upsilon-2}[\Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}] \end{aligned}$$