Contents

1 Comments to slides

1.1 Comment to Distribution

Discrete random variable - **probability function** (pf)

$$f_Y(y) = P(Y = y)$$

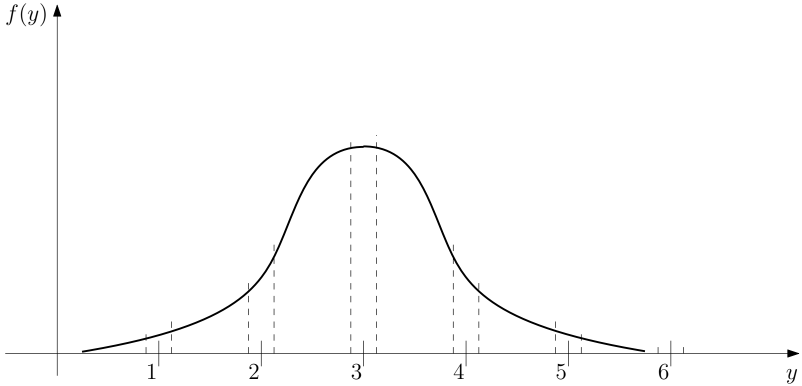
Continuous random variable - **probability density function** (pdf)

$$F_Y(y) = P(Y \leq y) \text{ distribution function}$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} \rightarrow F_Y(y) = \int_{-\infty}^{y} f_Y(t)\, dt$$

$\cdots$ probability of a single value is always zero !!!

Continuous random variable



Instead of values of y we will speak about their neighborhoods O_y. Then probabilities of values must be understood as probabilities of their neighborhoods. By "higher probability of y" we understand higher probability of values from its neighborhood.

1.2 Model of a false coin

$y_t \in \{1, 2\}$, $P(y_t = 1) = p$; $P(y_t = 2) = 1 - p$

Probability function

y_t	1	2
$f(y_t)$	p_1	p_2

where $p_1, p_2 \geq 0$ and $p_1 + p_2 = 1$.

Two coins

coin = 1,2

$$f(x|\text{coin})$$

coin	1	2
1	p_{11}	p_{12}
2	p_{21}	p_{22}

1.3 Controlled coin with memory

$u_t \in \{1,2\}$ - control variable

$$f\left(y_t|u_t, y_{t-1}\right)$$

$u_t,\ y_{t-1}$	$y_t = 1$	$y_t = 2$
1 1	$\Theta_{1\|11}$	$\Theta_{2\|11}$
1 2	$\Theta_{1\|12}$	$\Theta_{2\|12}$
2 1	$\Theta_{1\|21}$	$\Theta_{2\|21}$
2 2	$\Theta_{1\|22}$	$\Theta_{2\|22}$

where $1\,1 \to 1$; $1\,2 \to 2$; $2\,1 \to 3$; $2\,2 \to 4$ can be viewed as denotation of four coins (each with its own parameters). The choice of the coin is given by the choice of u_t and the last result y_{t-1}.

Two variables coded to one.

1.4 Bayes rule

Derivation

$$\begin{aligned} f\left(A,B|C\right) &= f\left(A|B,C\right)f\left(B|C\right) \\ &= f\left(B|A,C\right)f\left(A|C\right) \end{aligned}$$

$$f\left(A|B,C\right) = \frac{f\left(B|A,C\right)f\left(A|C\right)}{f\left(B|C\right)} \propto f\left(B|A,C\right)f\left(A|C\right)$$

Application to estimation

$$A = \Theta; \;\; B = y_t; \;\; C = \left\{u_t, d\left(t-1\right)\right\}$$

$$f\left(\Theta|\underbrace{y_t,\left\{u_t,d\left(t-1\right)\right\}}_{d(t)}\right) \propto f\left(y_t|\underbrace{\left\{u_t,d\left(t-1\right)\right\}}_{\psi_t},\Theta\right) f\left(\Theta|\underbrace{\left\{u_t,d\left(t-1\right)\right\}}_{d(t-1)}\right)$$

$$\underbrace{f\left(\Theta|d\left(t\right)\right)}_{\text{posterior}} \propto \underbrace{f\left(y_t|\psi_t,\Theta\right)}_{\text{model}}\underbrace{f\left(\Theta|d\left(t-1\right)\right)}_{\text{prior}}$$

1.5 Recursive estimation

Model $\quad f\left(y_t|\psi_t,\Theta\right) = a\exp\left(-ay_t\right) \cdots \quad \Theta = a,\ \psi_t = []$

For $t = 1, 2, \cdots$ the likelihood will be:

$t = 1$

$$L_1 = a\exp\left(-ay_1\right)$$

$t = 2$

$$L_2 = a\exp\left(-ay_1\right)\ a\exp\left(-ay_2\right) = a^2\exp\left(-a\left(y_1+y_2\right)\right)$$

For general time t

$$L_t = a\exp\left(-ay_t\right)\underbrace{a^{t-1}exp\left(-a\sum_{i=1}^{t-1}y_i\right)}_{L_{t-1}} = a^t\exp\left(-a\sum_{i=1}^{t}y_i\right)$$

Statistics

$$\kappa_t = t, \quad S_t = \sum_{i=1}^{t}y_i$$

Final likelihood

$$L_t = a^{\kappa_t} \exp\left(-aS_t\right) \cdots \text{ Gamma distribution}$$

Update of statistics

$$\kappa_t = \kappa_{t-1} + 1; \quad S_t = S_{t-1} + y_t$$

$$\text{with } \kappa_0 = 0,\, S_0 = 0$$

Point estimates

$$\frac{d}{da} L_t = \kappa_t a^{\kappa_t} \exp\left(-aS_t\right) - a^{\kappa_t} S_t \exp\left(-aS_t\right) = 0$$

$$\kappa_t = aS_t \;\rightarrow\; \hat{a}_t = \frac{\kappa_t}{S_t} = \frac{1}{\bar{y}}$$

Prior information

$$\kappa_0 \neq 0, \quad S_0 \neq 0$$

e.g. from t_0 prior data the average $\bar{y}_0$ is guessed (either from data or from an expert - t_0 is then understand as the strength of the information)

Prior

$$\kappa_0 = t_0$$

$$\frac{S_0}{t_0} = \bar{y}_0 \;\rightarrow\; S_0 = t_0 \bar{y}_0$$

1.6 Static normal model in estimation

$$y_t = k + e_t,\ e \sim N\left(0,1\right)$$

$$f\left(y_t|k\right) \propto \exp\left(\frac{1}{2}\left(y_t - k\right)^2\right) = \exp\left(\frac{1}{2}\left(y_t^2 - 2ky_t + k^2\right)\right)$$

$$\prod_{t=1}^{n} f\left(y_t|k\right) \propto \exp\left(\frac{1}{2}\sum_{t=1}^{n}\left(y_t^2 - 2ky_t + k^2\right)\right) =$$
$$= \exp\left(\frac{1}{2}\left\{\sum_{t=1}^{n} y_i^2 - 2k\sum_{t=1}^{n} y_t + nk^2\right\}\right)$$

$$R = \sum_{t=1}^{n} y_t^2, \quad S = \sum_{t=1}^{n} y_t, \quad T = n$$

$$f\left(k|d\left(n\right)\right) \propto \exp\left(\frac{1}{2}\left\{R - 2kS + Tk^2\right\}\right)$$

maximum

$$\frac{d}{dk}f\left(k|d\left(n\right)\right)=\exp\left(\frac{1}{2}\left\{R-2kS+Tk\right\}\right)\frac{1}{2}\left\{-2S+2Tk\right\}=0$$

$$S=Tk$$

$$k=\frac{S}{T}=\frac{\sum_{t=1}^{n}y_t}{n}=\bar{y}$$

1.7 Regression model in estimation

Model (normal)

$$f\left(y_t|\psi_t,\Theta\right)\propto r^{-0.5}\exp\left\{-\frac{1}{2r}\left(y_t-\psi_t'\theta\right)^2\right\}=(*)$$

$$y_t-\psi_t'\theta=-\left[-1,\ \theta'\right]\begin{bmatrix}y_t\\ \psi_t\end{bmatrix}$$

$$\left(y_t-\theta\psi_t'\right)^2=\left[-1,\ \theta'\right]\underbrace{\begin{bmatrix}y_t\\ \psi_t\end{bmatrix}\left[y_t,\ \psi_t'\right]}_{D_t}\begin{bmatrix}-1\\ \theta\end{bmatrix}=\left[-1,\ \theta'\right]D_t\begin{bmatrix}-1\\ \theta\end{bmatrix}$$

$$(*)=r^{-0.5}\exp\left\{-\frac{1}{2r}\left[-1,\ \theta'\right]D_t\begin{bmatrix}-1\\ \theta\end{bmatrix}\right\}$$

Prior (Gauss-inverse-Wishart)

$$f\left(\Theta|d\left(0\right)\right)\propto r^{-0.5\kappa_0}\exp\left\{-\frac{1}{2r}\left[-1,\ \theta'\right]V_0\begin{bmatrix}-1\\ \theta\end{bmatrix}\right\}$$

1.8 Coin in estimation

Let us estimate a false coin

$$f\left(y_t|p_1,p_2\right)$$

y_t	1	2
$f\left(y_t\right)$	p_1	p_2

where $p_1, p_2 \geq 0$, $p_1 + p_2 = 1$.

Statistics

$$\nu = [\nu_1, \nu_2]$$

Update

$$\nu_{y_t} = \nu_{y_t} + 1$$

counts the results $y_t = 1$ and $y_t = 2$.

1.9 Discrete model in estimation

Model (categorical) in a product form

$$f\left(y_t|\psi_t,\Theta\right)=\Theta_{y_t|\psi_t}=\prod_{i|j}\Theta_{i|j}^{\delta(i|j,y_t|\psi_t)}$$

where $\delta\left(i|j,y_t|\psi_t\right)=1$ for $i|j=y_t|\psi_t$ and 0 otherwise (Kronecker).

Prior (Dirichlet)

$$f\left(\Theta|d\left(0\right)\right)\propto\prod_{i|j}\Theta_{i|j}^{\nu_{i|j;0}}$$

where the statistics v is a matrix with the same dimensions as model.

1.10 Shift of regression vector

used in program for prediction with regression model

Regression vector at t

$$\psi_t = \left[\underbrace{u_t, y_{t-1}, u_{t-1}, y_{t-2}, u_{t-2}}_{***}, y_{t-3}, u_{t-3}, 1 \right]'$$

Regression vector at $t+1$

$$\psi_{t+1} = \left[\underbrace{u_{t+1}, y_t}_{\text{new}}, \underbrace{u_t, y_{t-1}, u_{t-1}, y_{t-2}, u_{t-2}}_{***}, 1 \right]'$$

1.11 Derivative of vector function of vector argument

Function

$$g(x) = \begin{bmatrix} g_1(x_1, x_2 \cdots x_n) \\ g_2(x_1, x_2 \cdots x_n) \\ \cdots \\ g_m(x_1, x_2 \cdots x_n) \end{bmatrix}$$

Derivative

$$\frac{dg}{dx} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

Example

$$g\left(x\right)=\left[\begin{array}{c}3x_1-x_1x_2^2\\x_1^3+5x_2\end{array}\right]$$

$$\frac{\partial g_1}{\partial x_1}=\frac{\partial\left(3x_1-x_1x_2^2\right)}{\partial x_1}=3-x_2^2$$

$$\frac{\partial g_1}{\partial x_2}=\frac{\partial\left(3x_1-x_1x_2^2\right)}{\partial x_2}=2x_1x_2$$

$$\frac{\partial g_2}{\partial x_1}=\frac{\partial\left(x_1^3+5x_2\right)}{\partial x_1}=3x_1^2$$

$$\frac{\partial g_2}{\partial x_2}=\frac{\partial\left(x_1^3+5x_2\right)}{\partial x_2}=5$$

$$\frac{dg}{dx}=\left[\begin{array}{cc}3-x_2^2 & 2x_1x_2\\3x_1 & 5\end{array}\right]$$

1.12 Kalman filter as noise filter

g_t is a smooth signal which we measure with noise as signal y_t. We would like to estimate g_t as a state variable x_t from the noisy signal y_t.

Model

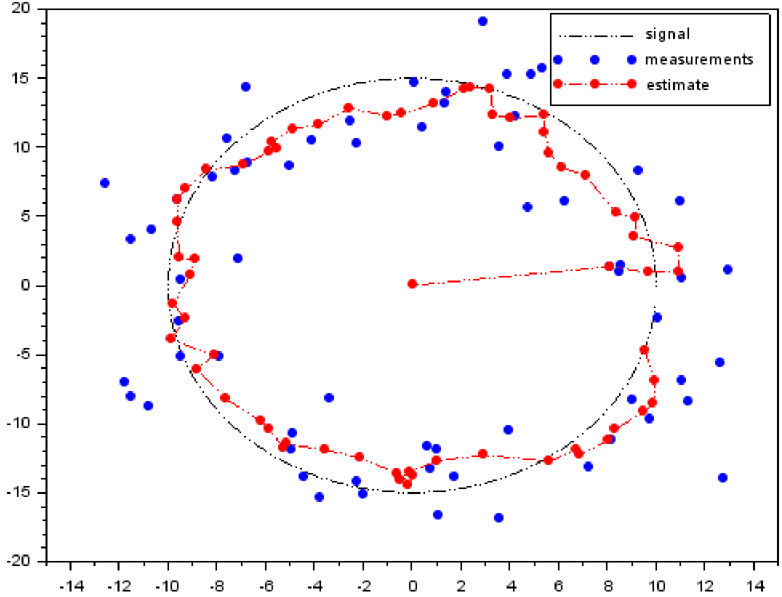
$$x_t = x_{t-1} + w_t$$
$$y_t = x_t + v_t$$

- variance of w_t determines the changes of the smooth signal
- variance of v_t indicates magnitude of the noise.

Program: T47statEst_Noise.sce

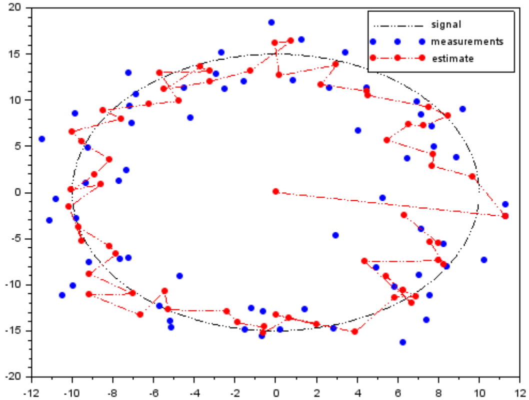
Noise filtration

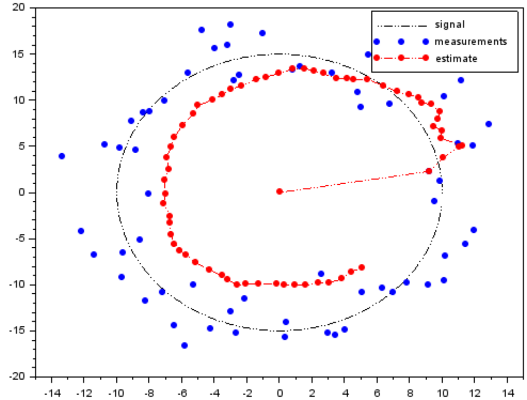
good



too big R_w

too small R_w





1.13 Kalman filter and unknown model parameters

Example

$$x_t = a x_{t-1} + u_t + w_t$$

$$y_t = x_t + v_t$$

New state $X_t = \begin{bmatrix} X_{1;t} \\ X_{2;t} \end{bmatrix} = \begin{bmatrix} x_t \\ a \end{bmatrix}$

New model

$$\begin{bmatrix} x_t \\ a \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ a \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \end{bmatrix} + \begin{bmatrix} w_{1;t} \\ w_{2;t} \end{bmatrix}$$

$$\begin{bmatrix} X_{1;t} \\ X_{2;t} \end{bmatrix} = \begin{bmatrix} X_{2;t-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1;t-1} \\ X_{2;t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \end{bmatrix} + W_t$$

$$X_{1;t} = X_{2;t-1} X_{1;t-1} + u_t + w_{1;t}$$

$$X_{2;t} = X_{2;t-1} + w_{2;t}$$

$\cdots$ must be linearized.

1.14 Derivation of control in pdf

$\cdots \;\; J_t = y_t^2 + \omega u_t^2; \quad \varphi_{N+1}^* = 0 \qquad\qquad \min_{u_{1:N}} E\left[\sum_{t=1}^{N} J_t | d\left(0\right)\right] =$

$$= \min_{u_{1:N}} E\left[\varphi_{N+1}^* + \sum_{t=1}^{N} J_t | d\left(0\right)\right] =$$

$$= \min_{u_{1:(N-1)}} E\left[\underbrace{\min_{u_N} \underbrace{E\left[\varphi_{N+1}^* + J_N | u_N, d\left(N-1\right)\right]}_{\varphi_N}}_{\varphi_N^*} + \sum_{t=1}^{N-1} J_t \middle| d\left(0\right)\right] =$$

$$= \min_{u_{1:(N-1)}} E\left[\min_{u_N} \varphi_N + \sum_{t=1}^{N-1} J_t | d\left(0\right)\right] = \min_{u_{1:(N-1)}} E\left[\varphi_N^* + \sum_{t=1}^{N-1} J_t | d\left(0\right)\right]$$

$\rightarrow$ Bellman equations (for $t = N, N-1, \cdots, 1$ do)

$$\varphi_t = E\left[\varphi_{t+1}^* + J_t | u_t, d\left(t-1\right)\right]$$

$$\varphi_t^* = \min_{u_t} \varphi_t$$

1.15 Control for regression model

In state form $\cdots$ regression model $\to$ state-space model (for 2nd order)

$$x_t = [y_t, u_t, y_{t-1}, u_{t-1}, y_{t-2}, u_{t-2}, 1]'$$

penalization

$$y_t^2 + \omega u_t^2 = x_t' \Omega x_t$$

where

$\Omega =$

	y_t	u_t	y_{t-1}	u_{t-1}	y_{t-2}	u_{t-2}	1
y_t	1						
u_t		ω					
y_{t-1}							
u_{t-1}							
y_{t-2}							
u_{t-2}							
1							

Derivation $\quad \varphi^*_{t+1} = x_t^{'} R_{t+1} x_t$ (assumption)

Bellman equations (**expectation** - model substitution)

$$E\left[x_t^{'} R_{t+1} x_t + x_t^{'} \Omega x_t | u_t, d\left(t-1\right)\right] = E\left[x_t^{'} U x_t | u_t, d\left(t-1\right)\right] =$$

$$= \left(M x_{t-1} + N u_t\right)' U \left(M x_{t-1} + N u_t\right) + \rho =$$

minimization

$$= x_{t-1}^{'} \underbrace{M'UM}_{C} x_{t-1} + 2u_t^{'} \underbrace{N'UM}_{B} x_{t-1} + u_t^{'} \underbrace{N'UN}_{A} u_t + \rho =$$

$$= u_t^{'} A u_t + 2u_t^{'} A \underbrace{A^{-1} B}_{S_t} x_{t-1} + x_{t-1}^{'} S_t^{'} A S_t x_{t-1} +$$

$$+ \underbrace{x_{t-1}^{'} C x_{t-1} - x_{t-1}^{'} S_t^{'} A S_t x_{t-1}}_{x_{t-1} R_t x_{t-1}} + \rho =$$

$$= \left(u_t + S_t x_{t-1}\right)' A \left(u_t + S_t x_{t-1}\right) + x_{t-1}^{'} R_t x_{t-1} + \rho$$

Optimal control $\quad u_t = -S_t x_{t-1}$ $\qquad$ Reminder $x_{t-1}^{'} R_t x_{t-1} + \rho$

Penalization for control increments and setpoint

$$J_t = (y_t - s_t)^2 + \omega u_t^2 + \lambda (u_t - u_{t-1})^2 =$$

$$= y_t^2 - 2y_t s_t + s_t^2 + \omega u_t^2 + \lambda u_t^2 - 2\lambda u_t u_{t-1} + \lambda u_{t-1}^2$$

$\Omega =$

	y_t	u_t	y_{t-1}	u_{t-1}	y_{t-2}	u_{t-2}	1
y_t	1						$-s_t$
u_t		$\omega+\lambda$		$-\lambda$			
y_{t-1}							
u_{t-1}		$-\lambda$		λ			
y_{t-2}							
u_{t-2}							
1	$-s_t$						s_t^2