Meeting 13:
More examples on the value of information (and with influence diagrams)

The decisive approach to statistical inference. Part I

Exercise 7.14

14. Suppose that a contractor must decide whether or not to build any speculative houses (houses for which he would have to find a buyer), and if so, how many. The houses that this contractor builds are sold for a price of $30,000, and they cost him $26,000 to build. Since the contractor cannot afford to have too much cash tied up at once, any houses that remain unsold three months after they are completed will have to be sold to a realtor for $25,000. The contractor's prior distribution for $\tilde{\theta}$, the number of houses that will be sold within three months of completion, is:

θ	$P(\tilde{\theta} = \theta)$
0	0.05
1	0.10
2	0.10
3	0.20
4	0.25
5	0.20
6	0.10

If the contractor's utility function is linear with respect to money, how many houses should he build? How much should he be willing to pay to find out for certain how many houses will be sold within three months?

The state of the world is how many houses, θ , that will be sold within tree months of completion. Possible values are 0, 1, 2, 3, 4, 5, 6.

The action to be taken is how many houses to be built,
i.e. a_k = "Build k houses" , k = 0,1,2,3,4,5,6 (the number cannot be higher than the maximum value of θ.

For each house built the payoff is 4 thousand dollars if it is sold and –1 thousand dollars if it has to be sold to a realtor.

The payoff function can then be written

$$R(a_k, \theta) = \begin{cases} 4 \cdot k & \text{if } k \leq \theta \\ 4 \cdot \theta - 1 \cdot (k - \theta) = 5 \cdot \theta - k & \text{if } k > \theta \end{cases}$$

The payoff table then becomes

a_k \ θ	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	–1	4	4	4	4	4	4
2	–2	3	8	8	8	8	8
3	–3	2	7	12	12	12	12
4	–4	1	6	11	16	16	16
5	–5	0	5	10	15	20	20
6	–6	–1	4	9	14	19	24

The (prior) probability distribution of $\tilde{\theta}$ is

θ	0	1	2	3	4	5	6
$P(\tilde{\theta}) = \theta$	0.05	0.10	0.10	0.20	0.25	0.20	0.10

The (prior) expected payoffs for each action are

a_k	$ER(a_k)$
0	0
1	(–1)·0.05 + 4·0.10 + 4·0.10 + 4·0.20 + 4·0.25 + 4·0.20 + 4·0.10 = 3.75
2	(–2)·0.05 + 3·0.10 + 8·0.10 + 8·0.20 + 8·0.25 + 8·0.20 + 8·0.10 = 7
3	(–3)·0.05 + 2·0.10 + 7·0.10 + 12·0.20 + 12·0.25 + 12·0.20 + 12·0.10 = 9.75
4	(–4)·0.05 + 1·0.10 + 6·0.10 + 11·0.20 + 16·0.25 + 16·0.20 + 16·0.10 = 11.5
5	(–5)·0.05 + 0·0.10 + 5·0.10 + 10·0.20 + 15·0.25 + 20·0.20 + 20·0.10 = **12**
6	(–6)·0.05 + (–1)·0.10 + 4·0.10 + 9·0.20 + 14·0.25 + 19·0.20 + 24·0.10 = 11.5

⇒ The optimal action according to the *ER*-criterion is to build 5 houses.

Losses for a_5 ($L(a_k, \theta) = \max_i\big(R(a_i, \theta)\big) - R(a_k, \theta)$)

θ	0	1	2	3	4	5	6
$\max_k\big(R(a_k, \theta)\big)$	0	4	8	12	16	20	24
$R(a_5, \theta)$	–5	0	5	10	15	20	20
$L(a_5, \theta)$	5	4	3	2	1	0	4

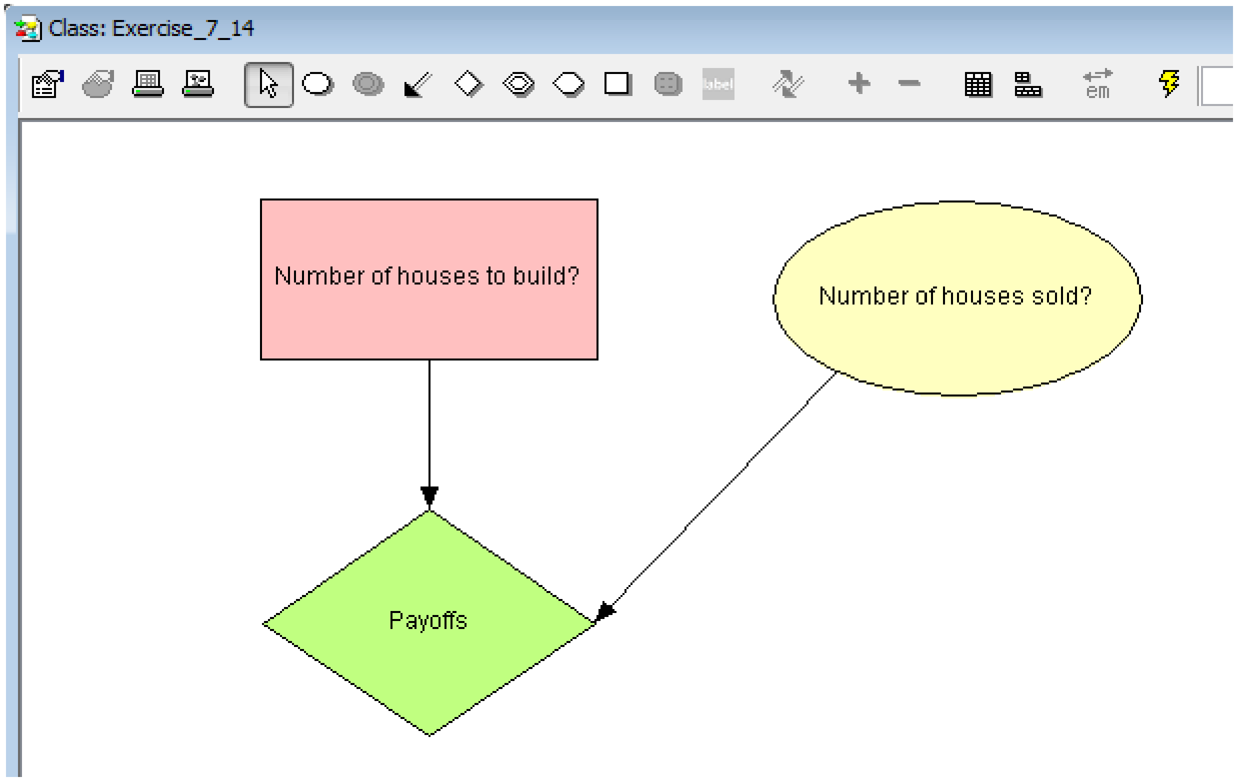
⇒ The expected loss for action a_5 is

$5 \cdot 0.05 + 4 \cdot 0.10 + 3 \cdot 0.10 + 2 \cdot 0.20 + 3 \cdot 0.25 + 0 \cdot 0.20 + 4 \cdot 0.10 = 2$

Since $EVPI = EL(a_{\text{opt}})$ the contractor should be willing to pay 2 thousand dollars to find out for certain how many houses will be sold.

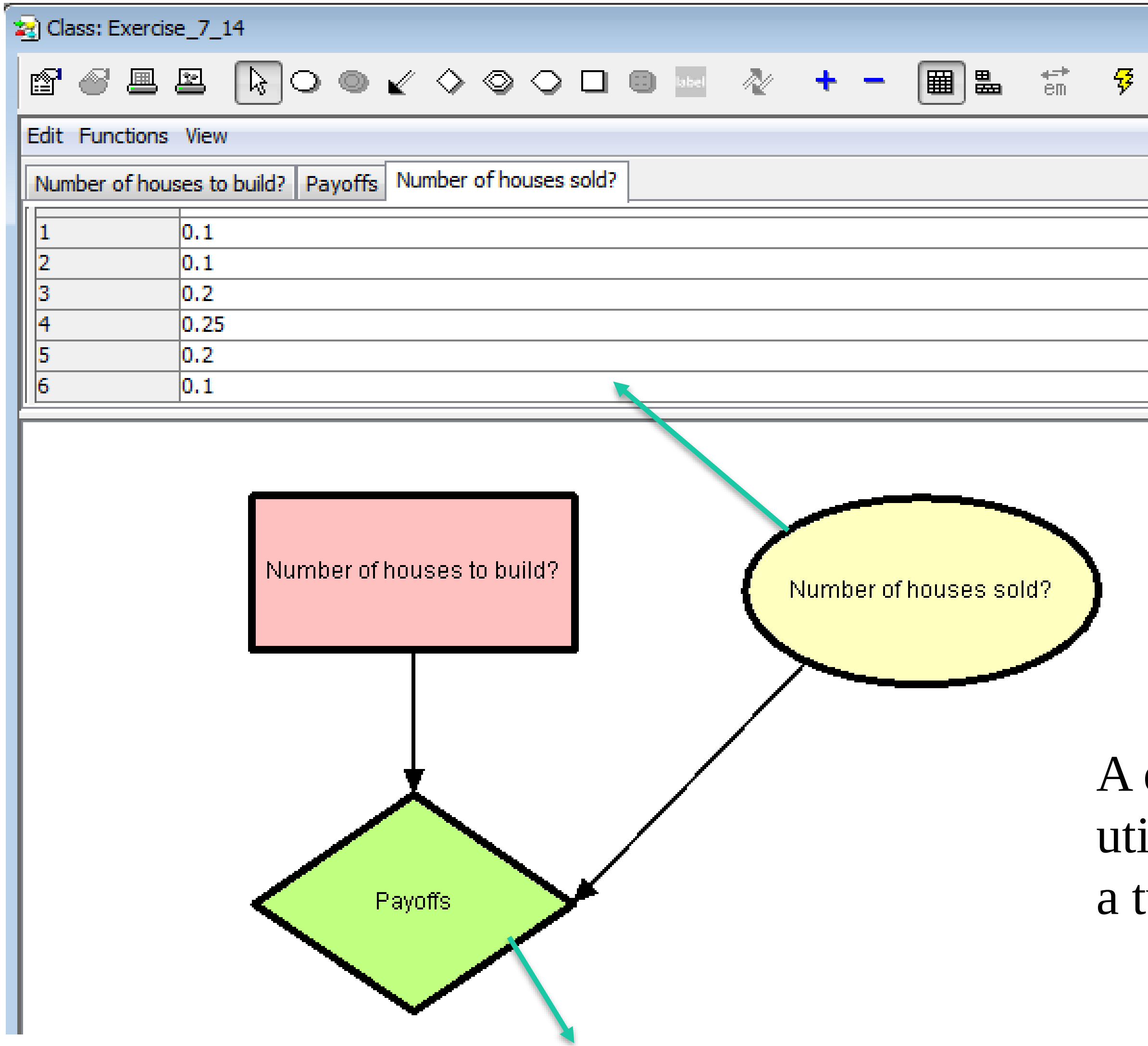
With Hugin:

Influence diagram:



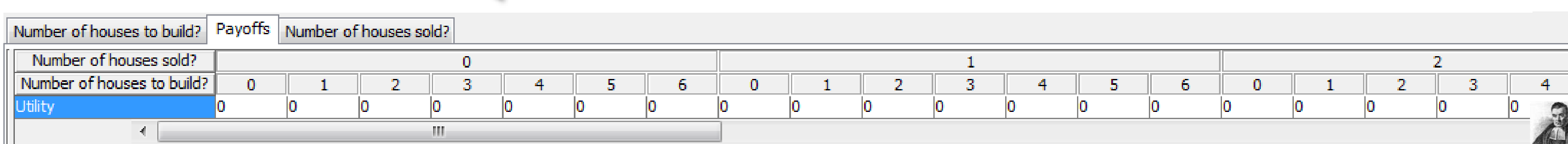
Note that no evidence node with observed data is present. The inference is from the prior distribution of θ.

Tables:



1	0.1
2	0.1
3	0.2
4	0.25
5	0.2
6	0.1

A drawback here is that the utility table cannot be viewed as a two-way table.



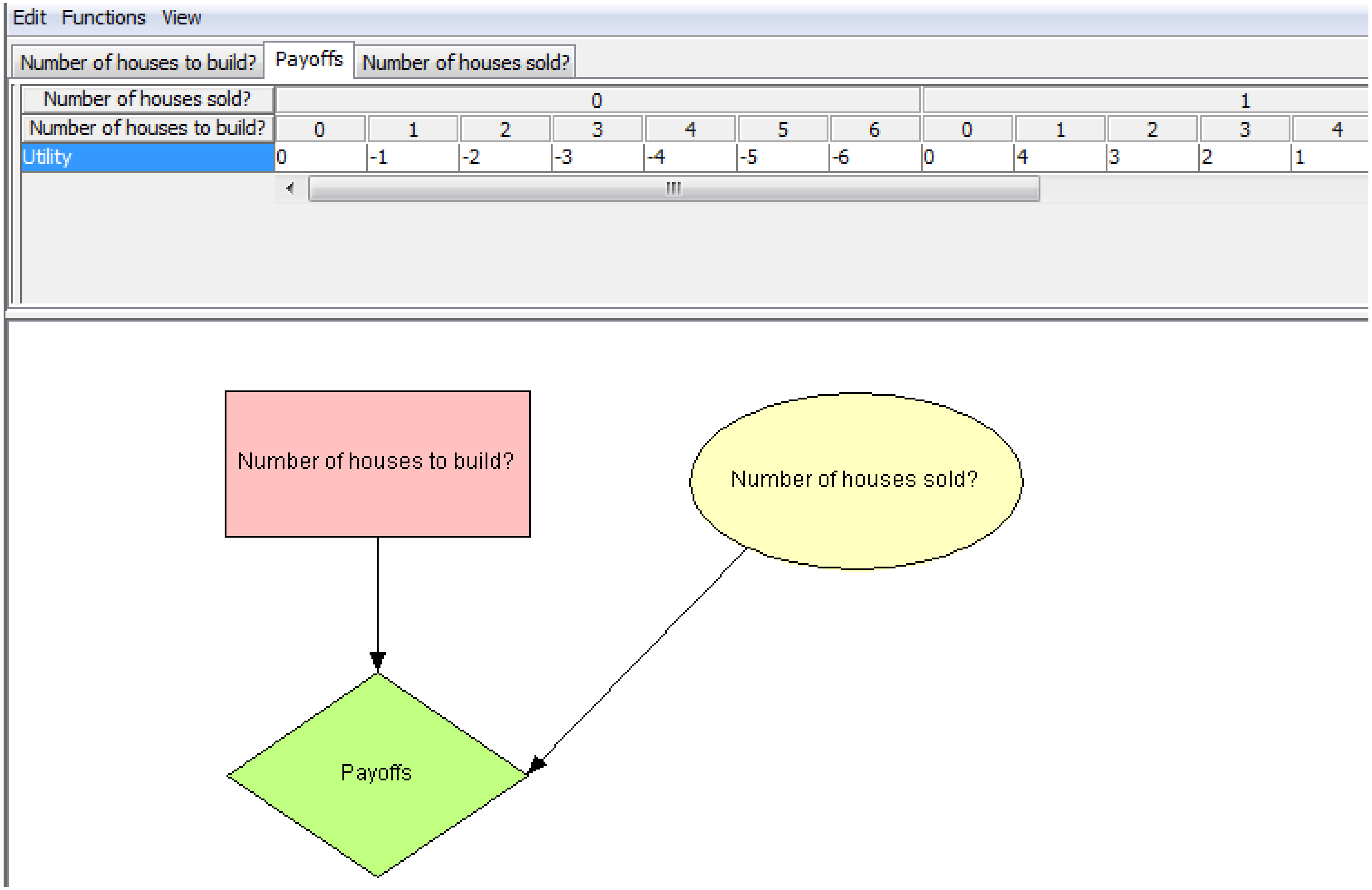
Number of houses sold?	0							1							2				
Number of houses to build?	0	1	2	3	4	5	6	0	1	2	3	4	5	6	0	1	2	3	4
Utility	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Number of houses to build? | Payoffs | Number of houses sold?

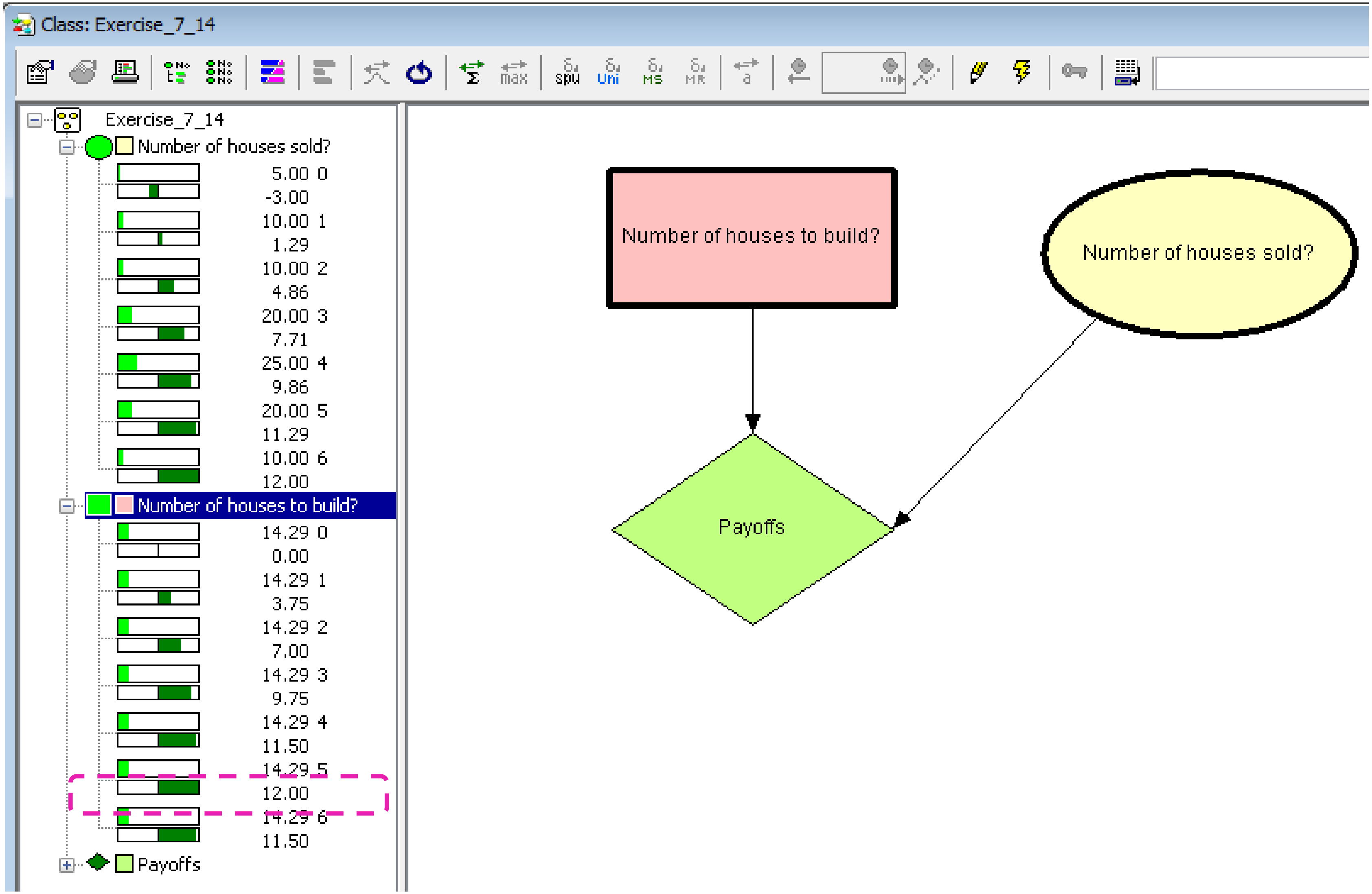
Number of houses sold?	0							1							2				
Number of houses to build?	0	1	2	3	4	5	6	0	1	2	3	4	5	6	0	1	2	3	4
Utility	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Utilities (here payoffs) from the payoff table shall be entered column wise.

a_k \ θ	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	–1	4	4	4	4	4	4
2	–2	3	8	8	8	8	8
3	–3	2	7	12	12	12	12
4	–4	1	6	11	16	16	16
5	–5	0	5	10	15	20	20
6	–6	–1	4	9	14	19	24



Run the network (flash icon).


Class: Exercise_7_14
Exercise_7_14
Number of houses sold?
5.00 0
-3.00
10.00 1
1.29
10.00 2
4.86
20.00 3
7.71
25.00 4
9.86
20.00 5
11.29
10.00 6
12.00
Number of houses to build?
14.29 0
0.00
14.29 1
3.75
14.29 2
7.00
14.29 3
9.75
14.29 4
11.50
14.29 5
12.00
14.29 6
11.50
Payoffs
Number of houses to build?
Number of houses sold?
Payoffs

We could also as utilities enter the losses with negative sign.

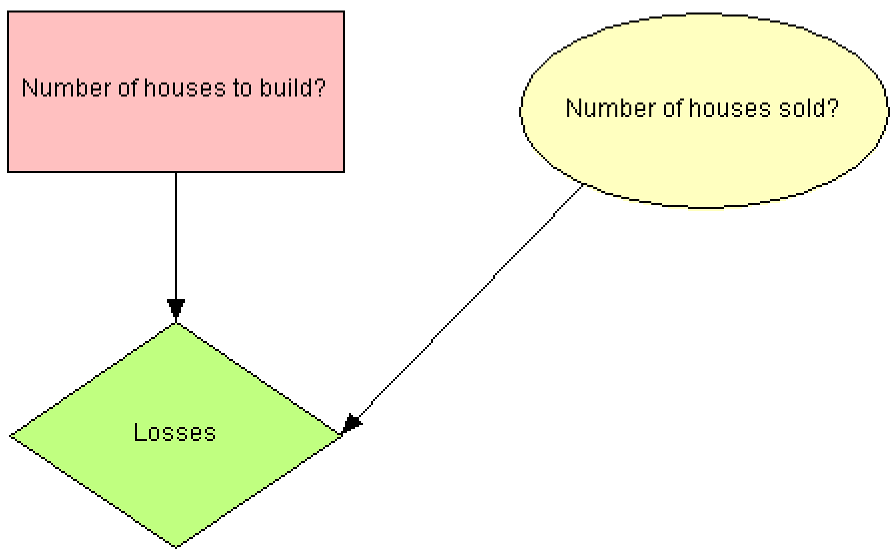
$$L(a_k, \theta) = \max_i\big(R(a_i, \theta)\big) - R(a_k, \theta)$$

a_k \ θ	0	1	2	3	4	5	6
0	0 – 0 = 0	4 – 0 = 4	8 – 0 = 8	12 – 0 = 12	16 – 0 = 16	20 – 0 = 20	24 – 0 = 24
1	0 – (–1) = 1	4 – 4 = 0	8 – 4 = 4	12 – 4 = 8	16 – 4 = 12	20 – 4 = 16	24 – 4 = 20
2	0 – (–2) = 2	4 – 3 = 1	8 – 8 = 0	12 – 8 = 4	16 – 8 = 8	20 – 8 = 12	24 – 8 = 16
3	0 – (–3) = 3	4 – 2 = 2	8 – 7 = 1	12 – 12 = 0	16 – 12 = 4	20 – 12 = 8	24 – 12 = 12
4	0 – (–4) = 4	4 – 1 = 3	8 – 6 = 2	12 – 11 = 1	16 – 16 = 0	20 – 16 = 4	24 – 16 = 8
5	0 – (–5) = 5	4 – 0 = 4	8 – 5 = 3	12 – 10 = 2	16 – 15 = 1	20 – 20 = 0	24 – 20 = 4
6	0 – (–6) = 6	4 – (–1) = 5	8 – 4 = 4	12 – 9 = 3	16 – 14 = 2	20 – 19 = 1	24 – 24 = 0
$\max_k\big(R(a_k, \theta)\big)$	0	4	8	12	16	20	24

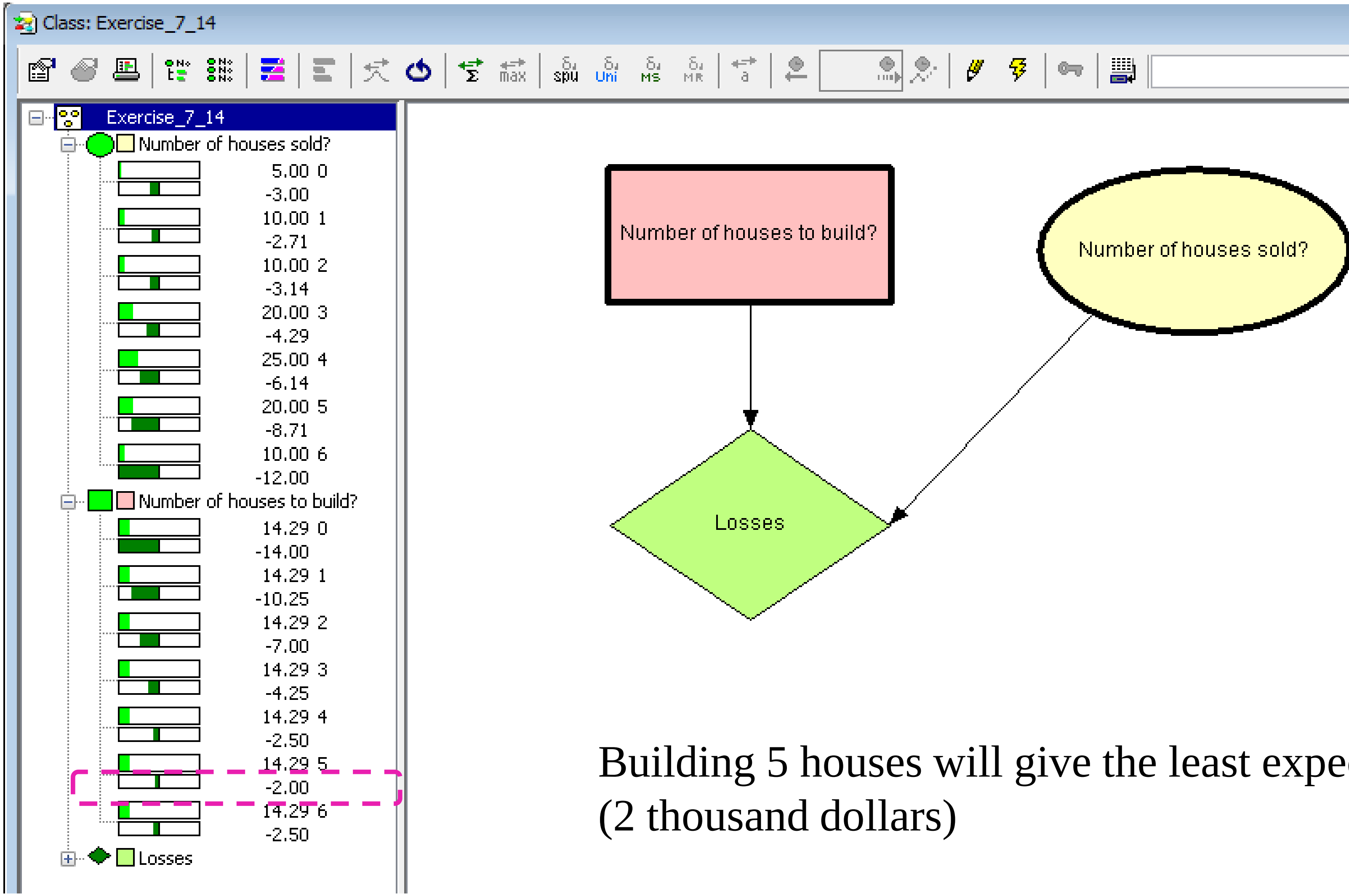
Edit Functions View

Number of houses to build? | Losses | Number of houses sold?

Number of houses sold?	0							1							
Number of houses to build?	0	1	2	3	4	5	6	0	1	2	3	4	5	6	0
Utility	0	-1	-2	-3	-4	-5	-6	-4	0	-1	-2	-3	-4	-5	-8



…and run.



Building 5 houses will give the least expected loss (2 thousand dollars)

…and we can at the same time answer the second question (*EVPI* = 2 000 $)

The Hugin software (like several other software) is limited when it comes to continuous probability distributions.

Commercial Hugin licences can handle normally distributed nodes but not other continuous probability distributions.

This “drawback” is due to that the internal probability calculations within the Hugin engine are exact, and therefore conjugate families with analytically deduced normalisation constants are the only feasible ones.

Many problems may however get acceptable solutions by discretising the probability density functions.

A lot of Bayesian inference can be achieved by using MCMC, but the demonstration of the solution becomes much less transparent.

Some research have been made for implementing algorithms based on numerical integration (and not MCMC) to obtain approximations to exact probability calculus in Bayesian networks: “Simonsson I. (2018). *Exact inference in Bayesian networks and applications in forensic statistics*. Doktorsavhandlingar vid Chalmers tekniska högskola – No. 4499. Chalmers University of Technology, Sweden.”

Exercise 7.15

15. A hot-dog vendor at a football game must decide in advance how many hot dogs to order. He makes a profit of $0.10 on each hot dog that is sold, and he suffers a $0.20 loss on hot dogs that are unsold. If his distribution of the number of hot dogs that will be demanded at the football game is a normal distribution with mean 10,000 and standard deviation 2000, how many hot dogs should he order? How much is it worth to the vendor to know in advance exactly how many hot dogs will be demanded?

Let θ = Demand in no. of hot dogs

a_k = Order k hot dogs (action)

The payoff function is (cf. *Exercise 7.14*):

$$R(a_k, \theta) = \begin{cases} 0.10 \cdot k & \text{if } k \le \theta \\ 0.10 \cdot \theta - 0.20 \cdot (k - \theta) = 0.3 \cdot \theta - 0.2 \cdot k & \text{if } k > \theta \end{cases}$$

The prior distribution of $\tilde{\theta}$ is $N(\mu = 10000, \sigma = 2000)$

$\Longrightarrow$ The pdf of $\tilde{\theta}$ is $f_{\tilde{\theta}}(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}$ and the cdf of $\tilde{\theta}$ is

$$F_{\tilde{\theta}}(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{x=\infty}^{\theta} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

The expected payoff with action a_k is

$$E\big(R(a_k,\theta)\big) = \int_{\theta=-\infty}^{\infty} R(a_k,\theta)\cdot f_{\tilde{\theta}}(\theta)d\theta =$$

$$= \int_{\theta=-\infty}^{k} (0.3\cdot\theta - 0.2\cdot k)\cdot f_{\tilde{\theta}}(\theta)d\theta + \int_{\theta=k}^{\infty} 0.1\cdot k\cdot f_{\tilde{\theta}}(\theta)d\theta =$$

$$= 0.3\cdot \int_{\theta=-\infty}^{k} \theta\cdot f_{\tilde{\theta}}(\theta)d\theta - 0.2\cdot k\cdot \int_{\theta=-\infty}^{k} f_{\tilde{\theta}}(\theta)d\theta + 0.1\cdot k\cdot \int_{\theta=k}^{\infty} f_{\tilde{\theta}}(\theta)d\theta =$$

$$= 0.3\cdot \int_{\theta=-\infty}^{k} \theta\cdot f_{\tilde{\theta}}(\theta)d\theta - 0.2\cdot k\cdot F_{\tilde{\theta}}(k) + 0.1\cdot k\cdot \Big(1 - F_{\tilde{\theta}}(k)\Big) =$$

$$= 0.3\cdot \int_{\theta=-\infty}^{k} \theta\cdot f_{\tilde{\theta}}(\theta)d\theta - 0.3\cdot k\cdot F_{\tilde{\theta}}(k) + 0.1\cdot k$$

$$\int_{\theta=-\infty}^{k} \theta \cdot f_{\tilde{\theta}}(\theta) d\theta = \int_{\theta=-\infty}^{k} \theta \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta = \langle \text{Form Gaussian integrals} \rangle =$$

$$= \int_{\theta=-\infty}^{k} \left(-\sigma^2 \cdot \left(-\frac{\theta-\mu}{\sigma^2}\right) + \mu\right) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta =$$

$$= -\sigma^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{\theta=-\infty}^{k} \left(-\frac{\theta-\mu}{\sigma^2}\right) \cdot e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta + \mu \cdot \int_{\theta=-\infty}^{k} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta =$$

$$= -\sigma^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \left[e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}\right]_{\theta=-\infty}^{k} + \mu \cdot F_{\tilde{\theta}}(k) = -\sigma^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \left(e^{-\frac{(k-\mu)^2}{2\sigma^2}} - 0\right) + \mu \cdot F_{\tilde{\theta}}(k) =$$

$$= -\sigma^2 \cdot f_{\tilde{\theta}}(k) + \mu \cdot F_{\tilde{\theta}}(k)$$

$$\Longrightarrow E\big(R(a_k, \theta)\big) = 0.3 \cdot \left(-\sigma^2 \cdot f_{\tilde{\theta}}(k) + \mu \cdot F_{\tilde{\theta}}(k)\right) - 0.3 \cdot k \cdot F_{\tilde{\theta}}(k) + 0.1 \cdot k =$$

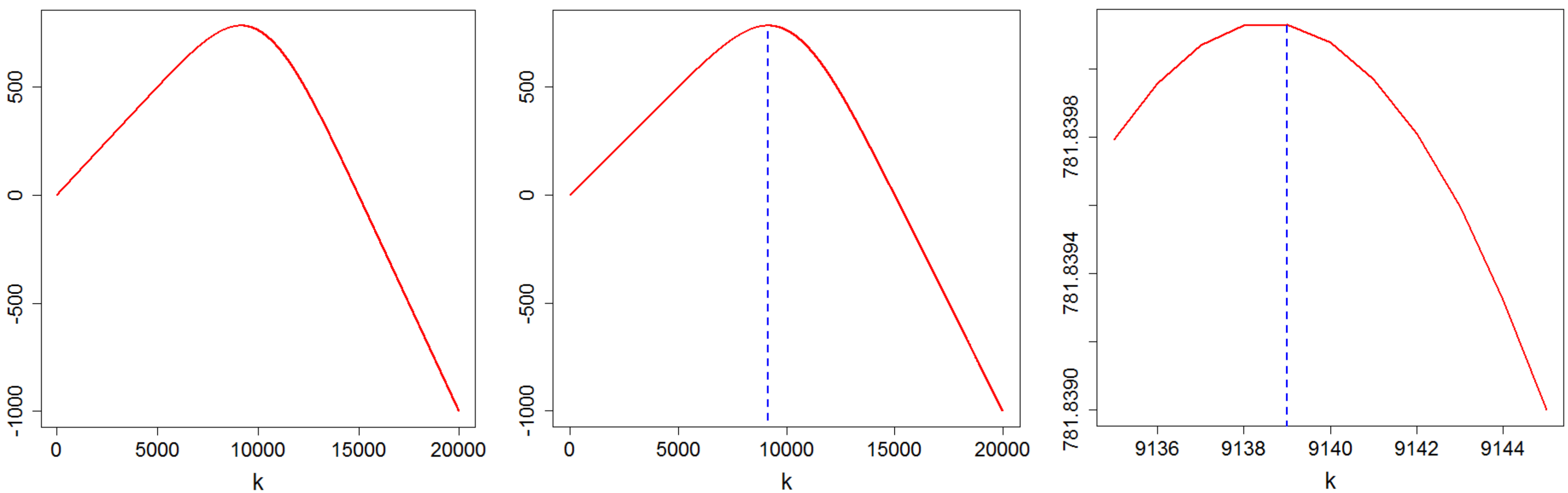
$$= 0.3 \cdot (\mu - k) \cdot F_{\tilde{\theta}}(k) - 0.3 \cdot \sigma^2 \cdot f_{\tilde{\theta}}(k) + 0.1 \cdot k$$

Hence, the optimal action with respect to the *ER*-criterion will be obtained from

$$\max_k \{0.3 \cdot (\mu - k) \cdot F_{\tilde{\theta}}(k) - 0.3 \cdot \sigma^2 \cdot f_{\tilde{\theta}}(k) + 0.1 \cdot k \}$$

A bit tricky to find the maximum from differential calculus.

Search maximum by graphing and grid-searching. Note that k must be an integer.



The optimal action is to order 9139 hot dogs.

$EVPI = EL(a_{\text{opt}})$

$$L(a_k, \theta) = \max_j \left(R\left(a_j, \theta\right)\right) - R(a_k, \theta)$$

$$R(a_k, \theta) = \begin{cases} 0.1 \cdot k & \text{if } k \leq \theta \\ 0.3 \cdot \theta - 0.2 \cdot k & \text{if } k > \theta \end{cases}$$

$\Longrightarrow \max_j \left(R\left(a_j, \theta\right)\right)$ is obtained when $a_j = \theta$

$$\Longrightarrow L(a_k, \theta) = R(\theta, \theta) - R(a_k, \theta)$$

$$L(a_k, \theta) = \begin{cases} 0.1 \cdot (\theta - k) & \text{if } k \leq \theta \\ 0.2 \cdot (k - \theta) & \text{if } k > \theta \end{cases}$$

Could be intuitively found

$$E\left(L(a_k)\right) = \int_{\theta=-\infty}^{\infty} L(a_k, \theta) \cdot f_{\tilde{\theta}}(\theta) d\theta =$$

$$= \int_{\theta=-\infty}^{k} 0.2 \cdot (k - \theta) \cdot f_{\tilde{\theta}}(\theta) d\theta + \int_{\theta=k}^{\infty} 0.1 \cdot (\theta - k) \cdot f_{\tilde{\theta}}(\theta) d\theta =$$

$$= 0.2 \cdot k \int_{\theta=-\infty}^{k} f_{\tilde{\theta}}(\theta)d\theta - 0.2 \cdot \int_{\theta=-\infty}^{k} \theta \cdot f_{\tilde{\theta}}(\theta)d\theta +$$

$$+0.1 \cdot \int_{\theta=k}^{\infty} \theta \cdot f_{\tilde{\theta}}(\theta)d\theta - 0.1 \cdot k \int_{\theta=k}^{\infty} f_{\tilde{\theta}}(\theta)d\theta = \langle \text{from previous calculations} \rangle =$$

$$= 0.2 \cdot k \cdot F_{\tilde{\theta}}(k) - 0.2 \cdot \left(-\sigma^2 \cdot f_{\tilde{\theta}}(k) + \mu \cdot F_{\tilde{\theta}}(k)\right) + 0.1 \cdot \int_{\theta=k}^{\infty} \theta \cdot f_{\tilde{\theta}}(\theta)d\theta$$

$$-0.1 \cdot k \cdot \left(1 - F_{\tilde{\theta}}(k)\right)$$

$$\int_{\theta=k}^{\infty} \theta \cdot f_{\tilde{\theta}}(\theta)d\theta = \langle \text{Analogous to previous calculations} \rangle =$$

$$= -\sigma^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \left[e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}\right]_{\theta=k}^{\infty} + \mu \cdot \left(1 - F_{\tilde{\theta}}(k)\right) = -\sigma^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \left(0 - e^{-\frac{(k-\mu)^2}{2\sigma^2}}\right) +$$

$$+\mu \cdot \left(1 - F_{\tilde{\theta}}(k)\right) = \sigma^2 \cdot f_{\tilde{\theta}}(k) + \mu \cdot \left(1 - F_{\tilde{\theta}}(k)\right)$$

$$\Longrightarrow E\big(L(a_k)\big) = 0.2 \cdot k \cdot F_{\tilde{\theta}}(k) - 0.2 \cdot \left(-\sigma^2 \cdot f_{\tilde{\theta}}(k) + \mu \cdot F_{\tilde{\theta}}(k)\right) +$$

$$+0.1 \cdot \left(\sigma^2 \cdot f_{\tilde{\theta}}(k) + \mu \cdot \left(1 - F_{\tilde{\theta}}(k)\right)\right) - 0.1 \cdot k \cdot \left(1 - F_{\tilde{\theta}}(k)\right) =$$

$$= 0.2 \cdot k \cdot F_{\tilde{\theta}}(k) - 0.2 \cdot \left(-\sigma^2 \cdot f_{\tilde{\theta}}(k)\right) - 0.2 \cdot \mu \cdot F_{\tilde{\theta}}(k) + 0.1 \cdot \left(\sigma^2 \cdot f_{\tilde{\theta}}(k)\right) + 0.1 \cdot \mu$$

$$-0.1 \cdot \mu \cdot F_{\tilde{\theta}}(k) - 0.1 \cdot k + 0.1 \cdot k \cdot F_{\tilde{\theta}}(k) =$$

$$= 0.3 \cdot k \cdot F_{\tilde{\theta}}(k) - 0.3 \cdot \mu \cdot F_{\tilde{\theta}}(k) + 0.3 \cdot \sigma^2 \cdot f_{\tilde{\theta}}(k) + 0.1 \cdot \mu - 0.1 \cdot k =$$

$$= 0.3 \cdot (k - \mu) \cdot F_{\tilde{\theta}}(k) + 0.3 \cdot \sigma^2 \cdot f_{\tilde{\theta}}(k) + 0.1 \cdot (\mu - k)$$

$$\Longrightarrow EVPI = E\left(L\left(a_{\text{opt}}\right)\right) = E\big(L(a_{9139})\big) =$$

$$= 0.3 \cdot (9139 - 10000) \cdot F_{\tilde{\theta}}(9139) + 0.3 \cdot 2000^2 \cdot f_{\tilde{\theta}}(9139) + 0.1 \cdot (10000 - 9139)$$

$$= -258.3 \cdot F_Z\left(\frac{9139 - 10000}{2000}\right) + 0.3 \cdot 2000^2 \cdot \frac{1}{2000} \cdot f_Z\left(\frac{9139 - 10000}{2000}\right) + 172.2 \approx$$

≈ 218 (dollar)

Alternative means of calculation:

$$E\big(L(a_k,\theta)\big) = E\big(R(\theta,\theta)\big) - E\big(R(a_k,\theta)\big) =$$

$$= E\big(R(\theta,\theta)\big) - 0.3 \cdot (\mu - k) \cdot F_{\tilde{\theta}}(k) + 0.3 \cdot \sigma^2 \cdot f_{\tilde{\theta}}(k) - 0.1 \cdot k$$

$$E\big(R(\theta,\theta)\big) = \int_{\theta=-\infty}^{\infty} R(\theta,\theta) \cdot f_{\tilde{\theta}}(\theta) d\theta =$$

$$= \int_{\theta=-\infty}^{k} (0.3 \cdot \theta - 0.2 \cdot \theta) \cdot f_{\tilde{\theta}}(\theta) d\theta + \int_{\theta=k}^{\infty} 0.1 \cdot \theta \cdot f_{\tilde{\theta}}(\theta) d\theta =$$

$$= \int_{\theta=-\infty}^{k} 0.1 \cdot \theta \cdot f_{\tilde{\theta}}(\theta) d\theta + \int_{\theta=k}^{\infty} 0.1 \cdot \theta \cdot f_{\tilde{\theta}}(\theta) d\theta = \int_{\theta=-\infty}^{\infty} 0.1 \cdot \theta \cdot f_{\tilde{\theta}}(\theta) d\theta =$$

$$= 0.1 \cdot E\big(\tilde{\theta}\big) = 1000$$

$$\Rightarrow E\Big(L\big(a_{\text{opt}}\big)\Big) = E\big(L(a_{9139})\big) =$$

$$= 1000 - 0.3 \cdot (10000 - 9139) \cdot F_{\tilde{\theta}}(9139) + 0.3 \cdot 2000^2 \cdot f_{\tilde{\theta}}(9139) - 0.1 \cdot 9139 =$$

$$\approx 218$$

The decisive approach to statistical inference, part I

Point estimation of an unknown parameter θ:

The <u>decision rule</u> is a point estimator (the functional form): $\delta(\widetilde{\boldsymbol{x}}) = \hat{\theta}(\widetilde{\boldsymbol{x}})$

The <u>action</u> is a particular point estimate. $\hat{\theta}_{obs} = \hat{\theta}(\boldsymbol{x})$

<u>State of nature</u> is the true value of θ.

The loss function is a measure of how far away the estimator is from θ:

$$L(\delta(\widetilde{\boldsymbol{x}}), \theta) = L\left(\hat{\theta}, \theta\right)$$

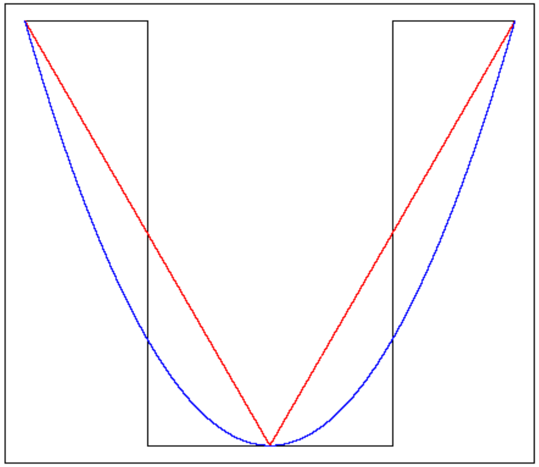
<u>Prior information</u> is quantified by the prior distribution (pdf/pmf) $f'(\theta)$.

<u>Data</u> is the random sample $\boldsymbol{x}$ from a distribution with (pdf/pmf) $f(\boldsymbol{x}|\theta)$.

Three simple loss functions (univariate case)

Zero-one loss:

$$L(\hat{\theta},\theta)=\begin{cases}0 & |\hat{\theta}-\theta|<m \\ k & |\hat{\theta}-\theta|\geq m\end{cases} \qquad k,m>0$$



Absolute error loss:

$$L(\hat{\theta},\theta)=k\cdot|\hat{\theta}-\theta| \qquad k>0$$

Quadratic (error) loss (or squared loss):

$$L(\hat{\theta},\theta)=k\cdot(\hat{\theta}-\theta)^2 \qquad k>0$$

Bayes estimators:

A Bayes estimator is the estimator that minimizes the expected posterior loss:

$$EL''\left(\hat{\theta}(\boldsymbol{x})\right) = \int_{\theta} L\left(\hat{\theta}(\boldsymbol{x}), \theta\right) \cdot f''(\theta|\boldsymbol{x}) d\theta$$

$$\Rightarrow \hat{\theta}_B(\boldsymbol{x}) = \min_{\delta}\left(\int_{\theta} L(\delta(\boldsymbol{x}), \theta) \cdot f''(\theta|\boldsymbol{x}) d\theta\right)$$

Minimization with respect to different loss functions will result in measures of location in the *posterior* distribution of θ.

Zero-one loss: $\hat{\theta}_B(\boldsymbol{x})$ is the posterior mode for θ given $\boldsymbol{x}$

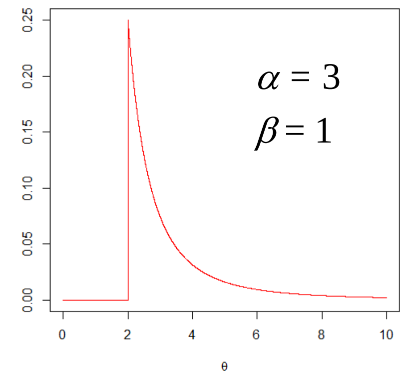
Absolute error loss: $\hat{\theta}_B(\boldsymbol{x})$ is the posterior median for θ given $\boldsymbol{x}$

Quadratic loss: $\hat{\theta}_B(\boldsymbol{x})$ is the posterior mean for θ given $\boldsymbol{x}: E(\theta|\boldsymbol{x})$

Example

Assume we have a sample $\boldsymbol{x} = (x_1, \ldots, x_n)$ from $U(0, \theta)$ and that a prior density for θ is the Pareto density

$$f'(\theta|\alpha,\beta) = (\alpha-1)\cdot\beta^{\alpha-1}\cdot\theta^{-\alpha} \quad ,\theta \geq 2; \alpha > 1; \beta > 0$$



What is the Bayes estimator of θ under quadratic loss?

The posterior distribution is also Pareto with

$x_{(n)} = max\{x_1, \ldots, x_n\}$

$$f''(\theta|n,\boldsymbol{x},\alpha,\beta) = (\alpha+n-1)\cdot\left(\max\{\beta,x_{(n)}\}\right)^{\alpha+n-1}\cdot\theta^{-(\alpha+n)} \quad ,\theta \geq \max\{\beta,x_{(n)}\}$$

$$\Rightarrow \tilde{\theta}_B = E(\tilde{\theta}|\boldsymbol{x}) = \int_{\theta=\max\{\beta,x_{(n)}\}}^{\infty} \theta\cdot(\alpha+n-1)\cdot\left(\max\{\beta,x_{(n)}\}\right)^{\alpha+n-1}\cdot\theta^{-(\alpha+n)}d\theta =$$

$$= (\alpha+n-1)\cdot\left(\max\{\beta,x_{(n)}\}\right)^{\alpha+n-1}\cdot\int_{\theta=\max\{\beta,x_{(n)}\}}^{\infty} \theta\cdot\theta^{-(\alpha+n-1)}d\theta =$$

$$= \frac{\alpha+n-1}{\alpha+n-2}\max\{\beta,x_{(n)}\}$$

Compare with $\hat{\theta}_{MLE} = x_{(n)}$

Minimax estimators:

Find the value of θ that maximizes the expected loss with respect to the sample values, i.e. that maximizes the risk over the set of estimators.

Then, the particular estimator that minimizes the risk for that value of θ is the minimax estimator.

$$\hat{\theta}_{\text{minimax}} = \underset{\delta}{\operatorname{argmin}} \left(\max_{\theta} D(\delta, \theta) \right)$$

Usually difficult to find minimax estimators, but there is one method to find it via a Bayes' estimator.

Theorem (actually a corollary of a theorem both presented in "Lehmann E.L. *Theory of point estimation*. Wiley, 1983")

If a Bayes' estimator has constant risk. [*i.e. not dependent on* θ] it is also a minimax estimator.

Example We wish to estimate the parameter p under quadratic loss in a binomial sampling model for sample size n. Hence we (will) have observed a random variable $\tilde{r}$ that is $Bi(n, p)$.

We (should) know that (with n fixed) the maximum-likelihood estimator of p is

$$\hat{p}_{MLE} = \frac{\tilde{r}}{n}$$

but can we find a minimax estimator under quadratic loss?

A Bayes' estimator under quadratic loss is the posterior mean of p, hence we need to specify a prior distribution – Natural to use the conjugate beta distribution with density function

$$f'(p|a,b) = \frac{p^{a-1}\cdot(1-p)^{b-1}}{B(a,b)} \quad 0 \le p \le 1$$

$$B(a,b) = \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)}$$

The posterior density function is then

$$f''(p|n,r,a,b) = \frac{p^{a+r-1}\cdot(1-p)^{b+n-r-1}}{B(a+r,b+n-r)}$$

and the Bayes' estimator is its mean:

$$\hat{p}_B = \hat{p}_B(\tilde{r}) = \frac{a+\tilde{r}}{a+r+b+n-r} = \frac{a+\tilde{r}}{a+b+n}$$

$$\hat{p}_B = \frac{a + \tilde{r}}{a + b + n}$$

The risk function for this estimator is

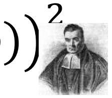
$$D(\hat{p}_B, p) = E_{\tilde{r}}\left((\hat{p}_B - p)^2\right) = \sum_{r=0}^{n} (\hat{p}_B - p)^2 \cdot \binom{n}{r} \cdot p^r \cdot (1-p)^r =$$

$$= \sum_{r=0}^{n} \left(\frac{a + r}{a + b + n} - p\right)^2 \cdot \binom{n}{r} \cdot p^r \cdot (1-p)^r =$$

$$= \sum_{r=0}^{n} \left[\left(\frac{a^2}{(a+b+n)^2} - \frac{2 \cdot a \cdot p}{a+b+n} + p^2\right) + \left(\frac{2 \cdot a}{(a+b+n)^2} - \frac{2 \cdot p}{a+b+n}\right) \cdot r +\right.$$

$$\left. + \frac{1}{(a+b+n)^2} \cdot r^2\right] \cdot \binom{n}{r} \cdot p^r \cdot (1-p)^r = \frac{a^2}{(a+b+n)^2} - \frac{2 \cdot a \cdot p}{a+b+n} + p^2 +$$

$$+ \left(\frac{2 \cdot a}{(a+b+n)^2} - \frac{2 \cdot p}{a+b+n}\right) \cdot \underbrace{\sum_{r=0}^{n} r \cdot \binom{n}{r} \cdot p^r \cdot (1-p)^r}_{E(\tilde{r}|p)} + \frac{1}{(a+b+n)^2} \cdot \underbrace{\sum_{r=0}^{n} r^2 \cdot \binom{n}{r} \cdot p^r \cdot (1-p)^r}_{E(\tilde{r}^2|p) = Var(\tilde{r}|p) + \left(E(\tilde{r}|p)\right)^2}$$



$$E(\tilde{r}|p) = n \cdot p$$

$$\hat{p}_B = \frac{a + \tilde{r}}{a + b + n}$$

$$Var(\tilde{r}|p) = n \cdot p \cdot (1 - p)$$

$$\Longrightarrow D(\hat{p}_B, p) = \frac{a^2}{(a+b+n)^2} - \frac{2 \cdot a \cdot p}{a+b+n} + p^2 + \left(\frac{2 \cdot a}{(a+b+n)^2} - \frac{2 \cdot p}{a+b+n}\right) \cdot n \cdot p +$$

$$+ \frac{1}{(a+b+n)^2} \cdot (n \cdot p \cdot (1-p) + (n \cdot p)^2) = \cdots =$$

$$= \frac{1}{(a+b+n)^2} \cdot \left(a^2 + \big(n - 2 \cdot a \cdot (a+b)\big) \cdot p + \big((a+b)^2 - n\big) \cdot p^2\right)$$

This risk function will be constant (for fixed *n*) if

$\big(n - 2 \cdot a \cdot (a+b)\big) = 0$ and $\big((a+b)^2 - n\big) = 0$

$$\Leftrightarrow a = b = \frac{\sqrt{n}}{2}$$

Hence, the estimator

$$\hat{p}_B = \frac{\frac{\sqrt{n}}{2} + \tilde{r}}{\frac{\sqrt{n}}{2} + \frac{\sqrt{n}}{2} + n} = \frac{\sqrt{n} + 2 \cdot \tilde{r}}{\sqrt{n} + n}$$

is a Bayes' estimator with constant risk, and according to the theorem above it is also a minimax estimator.

The value of the constant (but actually *n*-dependent) risk is

$$\frac{1}{\left(\frac{\sqrt{n}}{2} + \frac{\sqrt{n}}{2} + n\right)^2} \cdot \left(\frac{n}{4} + 0 \cdot p + 0 \cdot p^2\right) = \frac{n}{4 \cdot (\sqrt{n} + n)^2} = \frac{1}{4 \cdot (1 + \sqrt{n})^2}$$

> *Exercise:*
>
> Is $\hat{p}_B$ unbiased? What is the risk of the unbiased $\hat{p}_{MLE} = \tilde{r}/n$? For which range of n is the risk of $\hat{p}_B$ lower than the risk of $\hat{p}_{MLE}$?