1. $p(x|\theta)p(\theta)$ **is a distribution over:**

- ( ) θ
- (•) (x, θ)
- ( ) x

2. **Choose correct statements:**

1 / 1 point

- [x] $p(a, b \mid c) = p(a \mid b, c)p(b \mid c)$

✓ **Correct**

- [x] $p(a \mid b) = \int p(a \mid b, c)p(c)dc$, when b and c are independent

✓ **Correct**
$p(c \mid b) = p(c)$ when c and b are independent

- [ ] $p(a \mid b) = \int p(a \mid b, c)dc$

- [x] $p(a \mid b) = \int p(a, c \mid b)dc$

✓ **Correct**
The sum rule.

3. **Choose correct statements:**

- [ ] $p(a \mid b, c) = p(a \mid b)p(a \mid c)$ when b and c are independent

- [x] $p(a \mid b, c) = \frac{p(b \mid a, c)p(a \mid c)}{\int p(b \mid a', c)p(a' \mid c)da'}$

Correct

- [x] $p(a \mid b) = \frac{p(a, c \mid b)}{p(c \mid a, b)}$

Correct

Apply Bayes rule to $p(c \mid a,\, b)$

- [x] $p(a \mid b)p(b) + p(a \mid \bar{b})p(\bar{b}) = p(a)$, for binary b

Correct

The law of total probability

- [ ] $p(a \mid b) + p(a \mid \bar{b}) = p(a)$, for binary b

4. **Let joint probability over random variables a, b, c be $p(a, b, c) = p(a|b)p(b|c)p(c)$. Are random variables a and c independent?**

( ) Yes

(•) No

Correct

Let's marginalize joint probability by b and we get
$\int p(a, b, c)db = \int p(a|b)p(b|c)p(c)db = p(c) \int p(a|b)p(b|c)db$. Unfortunately, integral contain inside both a and c and it can't be decomposed into two integrals $\int f(a, b)db$ and $\int g(c, b)db$, so a and c is dependent.

5. **Let joint probability over random variables a, b, c, d be $p(a, b, c, d) = p(a|b)p(b)p(c|d)p(d)$. Are random variables a and c independent?**

- [x] Yes
- [ ] No

Correct

Let's marginalize joint probability by b and d, so we get
$\int p(a, b, c, d)\, db\, dd = \int p(a|b)p(b)p(c|d)p(d)\, db\, dd =$
$= \left(\int p(a|b)p(b)\, db\right) \left(\int p(c|d)p(d)\, dd\right)$. So we decomposed it into two integrals $\int f(a, b)db$ and $\int g(c, d)dd$, so a and c is independent.

6. **Recall the probabilistic regression setting. In the [lecture](), we have proved that solving the least-squares problem with L2 regularizer $L(w) = \sum_{i=1}^{n} (w^T x_i - y_i)^2 + C \sum_{i=1}^{N} w_i^2$ is equivalent to finding the MAP estimate for w with prior distribution $\mathcal{N}(w \mid 0, \gamma I)$. Let us now choose a prior distribution to be Laplace distribution: $p(w|0, b) = \frac{1}{(2b)^n} \prod_{i=1}^{n} \exp\left(-\frac{|w_i|}{b}\right)$ instead of Normal. Adding which of the following regularizers to the least-squares probem is equivalent to finding a MAP-estimate for such a model?**

- [ ] $\sum_{i=1}^{N} w_i^{\frac{1}{2}}$
- [x] $\sum_{i=1}^{N} |w_i|$
- [ ] $\sum_{i=1}^{N} \frac{1}{|w_i|}$
- [ ] $\sum_{i=1}^{N} w_i$

Correct

Let's get minus logarithm of Laplace distribution and we will get $C \sum_{i=1}^{n} |w_i| + D$, where C and D are some constants. We can forget about D constant because it's not important when we will minimize loss function.

7. **For linear regression with loss function $L(w) = \sum_{i=1}^{n}(w^T x_i - y_i)^2 + C\sum_{i=1}^{N} w_i^2$ prior distribution for weights w was Normal distribution $\mathcal{N}(w|0, \gamma I)$. Which prior distribution on weights is right for loss function $L(w) = \sum_{i=1}^{N}(w^T x_i - y_i)^2$ if each component of weights should be in some predefined range: $w_i \in [l, r]$?**

- [x] Uniform distribution with the same limits for each component.

$$p(w \mid a, b) = \begin{cases} \frac{1}{(b-a)^n}, \text{ if } \forall w_i \in [l, r] \\ 0, \text{ else} \end{cases}.$$

- [ ] Laplace distribution with zero mean and the same divergence for each component.

$$p(w \mid 0, b) = \frac{1}{(2b)^n} \prod_{i=1}^{n} \exp^{-\frac{|w_i|}{b}}.$$

- [ ] Gamma distribution with the same parameters for each component.

$$p(w \mid \alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)} \prod_i^n x_i^{\alpha-1} \exp^{-\beta w_i}$$

Correct

We can formulate question using equivalent loss function $L(w) = \sum_{i=1}^{N}(w^T x_i - y_i)^2 + \sum_{i=1}^{N} r(w_i)$ without limits on weights component but function $r(w_i) = \begin{cases} C, \text{ if } w_i \in [l, r] \\ +\infty, \text{ else} \end{cases}$. The same regularisation we will have if we find minus logarithm of Uniform distribution.

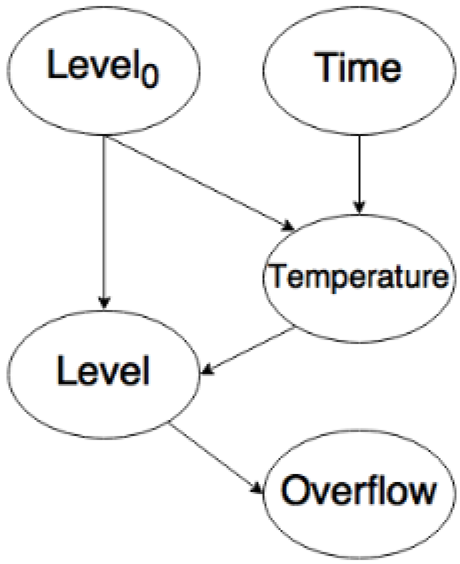
8. **For the remaining problems we will use the following story:**

You have a kettle that boils water. You pour water up to level L_0 and turn the kettle on. Over time, temperature $Temp$ starts to increase. At time T, level of water is L. Since water is boiling, water level slightly oscillates and so can be considered random. You also know that the height of a kettle is limited. If at some point water level exceeds this value, water will split on a table. We will denote this event as a binary random variable O (overflow). Our goal is to determine the maximum allowed initial water level L_{max} so that we can write it down in a kettle manual. Normally we would like to find L_{max} for which, for example, $P(O \mid L_0 = L_{max}) = 0.001$: if you pour this amount of water, overflow will occur with a fairly low probability.

In these tasks we will construct a Bayesian network and select probability distributions needed for the model.

Our first step is to choose the correct Bayesian network.

c)



9. **Write joint distribution for this situation.**

- $p(L_0|L, Temp)p(T|Temp)p(Temp|L)p(L|O)p(O)$
- $p(O|L)p(L|L_0, Temp)p(Temp|L_0, T)p(L_0)p(T)$
- $p(L|O)p(Temp|L_0, L)p(Temp|L_0, T)p(L_0)p(T)$
- $p(O|L)p(L|L_0, Temp)p(L_0, T|Temp)p(Temp)$

Correct

10. **Which distribution can you use for $p(L|L_0, Temp)$?**

- [x] a)