Exercises I

1. Assume an industrial production line in which two workers make products. Each of the workers may contaminate the product. Denote the events as E_i = "worker i contaminates the product", and A = "at least one of the workers contaminates the product". Make a table of truth values (T=true,F=false) for all combinations of the two events, with the corresponding truth value of the event A. Using indicator variables 1_{E_i} check that $1-(1-1_{E_1})(1-1_{E_2})$ is the indicator for event A.

2. Continue the previous problem by showing that $P(A) = P(E_1) + P(E_2) - P(E_1\&E_2)$.

3. Continue the previous problem: assume Mr X works from Mon to Fri, and Mr Y from Sat to Sun. Also assume that Mr X contaminates products with probability 0.01, and Mr Y with probability 0.05. A customer receives a contaminated product (but we don't know when it was produced). What is the probability it came from Mr X?

4. The joint probability for the three events at a swimming beach: R^+ ="it rains", W^+ ="it winds", K^+ ="ice cream kiosk open", is given in the table. ($R^- = \text{not}R^+$, etc.).

	R^+		R^-	
	W^+	W^-	W^+	W^-
K^+	1/40=0.025	2/40=0.05	5/40=0.125	9/40=0.225
K^-	10/40=0.25	8/40=0.2	3/40=0.075	2/40=0.05

What is the probability that the kiosk is closed, when it rains and winds? What is the probability that the kiosk is closed anyway? What is the probability that it rains anyway? What is the probability that it winds anyway? What is the probability that it rains and winds, when the kiosk is closed? Given that it does not rain, is the kiosk open independently of the wind? Show that you get the joint probability $P(K^+, R^+, W^+)$ by applying the product rule: $P(K^+ \mid R^+, W^+)P(R^+ \mid W^+)P(W^+)$.

5. In coin tossing, possible outcomes are $X = 0$ = 'heads', or $X = 1$ = 'tails'. Assume Bernoulli distribution $X \sim \text{Bern}(0.5)$. Depending on the outcome, either a balanced or loaded dice is chosen. The conditional distribution for the result from dice (Y = 1,2,...,6) is then categorical distribution $P(Y \mid X = 0) = \text{Cat}(\frac{1}{6}, \ldots, \frac{1}{6})$ or $P(Y \mid X = 1) = \text{Cat}(\frac{1}{10}, \ldots, \frac{1}{10}, \frac{1}{2})$. Write the joint distribution $P(X, Y)$ in a table. Calculate $P(X = 0 \mid Y = 6)$ from the table. Verify the result by solving $P(X = 0 \mid Y = 6)$ from Bayes theorem.

6. The following data represent measured pulses from 63 individuals, and repeated measurement after 1 minute. (http://www.statsci.org/data/oz/ms212.html). Open R (either in computer class, or download R from: http://www.r-project.org/), and copy paste these data on command prompt:

```
pulse1 <- c(86, 71, 90, 68, 71, 68, 74, 70, 78, 69, 77, 64, 83,
78, 88, 78, 80, 68, 62, 81, 58, 84, 104, 66, 84, 65, 80, 66, 104,
76, 70, 92, 63, 65, 76, 60, 80, 80, 76, 70, 80, 76, 83, 60, 80,
52, 74, 75, 72, 80, 74, 61, 76, 90, 64, 64, 82, 74, 119, 90, 86, 69, 75)
```

```
pulse2 <- c(88, 73,  88,  72,  77,  68,  76,  71,  82,  67,  73,
63,  79,  79,  86,  74,  76,  69,  59,  79,  58,  84,  92,  68,
90,  67,  80,  60,  96,  76,  68,  84,  65,  67,  74,  56,  72,
76,  72,  74,  80,  73,  84,  62,  82,  60,  72,  75, 68,  73,  72,
59,  80,  84,  68,  64,  87,  70, 120,  89,  84,  64,  68)
```

Plot a scatterplot of the data `plot(pulse1,pulse2)` to see the 2D empirical distribution. Plot also the histogram of the empirical marginal distributions, using narrow bins: `hist(pulse1,100)`, or smoothed density: `plot(density(pulse1))`, for both pulse1 and pulse2. From the empirical sample, plot the empirical distribution for $P(pulse2 \mid pulse1 = 80)$ using histograms and smoothed densities. Hint: you can get selected values from a vector by using indexing. For example, to select positive values from vector `v`, one could type `v[v>0]`. The resulting empirical distribution is based on only a few points, how many? This method could work better with larger data sets. How is Bayes theorem related to all this?