STAT 391
Homework 8
Out Thursday June 4, 2020
Due never

©Marina Meilă
mmp@stat.washington.edu

Some things I would like you to remember

0 – Start with the sample space

Many difficulties can be avoided by thinking of the sample space. Below is a very short illustration, and a very long one. In both cases, the problem is essentially solved once the set of outcomes is identified.

0.1 "I have two children, and one of them is a daughter." What is the probability that the other child is also a daughter? We assume that the sex of a child being a daugher is 0.5, independent of the sex of any other child.

0.2 See lecture notes.

1 – Write the likelihood *See also Homework 2* Data are sampled from a Poisson distribution $P(x) = \frac{\lambda^x}{x!}e^{-\lambda}$, $x \in \{0, 1, 2, \ldots\}$. By mistake, zeros and ones were counted together and you are given the following data

Value x	0 or 1	2	3	4	
# observed	8	6	4	2	Total $n = 20$

1.1. Write the expression of the log-likelihood l of the data as a function of λ.

1.2 Find the expression of the ML estimate of λ and the numerical value of λ^{ML} from the data above.

2 – The power of Bayes' rule

Nisqually River, Inc. sells books A,B,C on line. Each customer buys 0 or 1 copy of each title.

a. Mrs. Independence Day, the company's data mining expert, makes the assumptions that: i) a customer decides to buy a book independently of what other books (s)he buys and independently of other customers; ii) all customers buy according to the same *joint probability distribution* $\bar{P}_{ABC} = P_A P_B P_C$, with $P_A(1) = 0.6, P_B(1) = 0.3, P_C(1) = 0.4$. [For example, the probability that a customer buys A and B but not C is $\bar{P}_{ABC}(1, 1, 0) = 0.6 \times 0.3 \times (1 - 0.4)$.

Compute the probability that a customer buys all three books under $\bar{P}_{ABC}$, ie. compute

$$\bar{P}_{ABC}(1, 1, 1)$$

b. Mr. Mean Variance, her aide, insists that Mrs. Day's model is not correct. He assumes that ii) all customers buy according to the same probability distribution; but iii) Book C is bought independently from A, B but A, B are not independent of each other.

$$C \perp A, B \tag{1}$$
$$A \not\perp B \tag{2}$$

Denote Mr. Variance's model by $\tilde{P}_{ABC} = \tilde{P}_{AB}\tilde{P}_C$ with $\tilde{P}_C(1) = 0.4$, and $\tilde{P}_{AB}$:

	A = 0	A = 1
B = 0	0.2	0.5
1	0.2	0.1

[For example, $\tilde{P}_{ABC}(1, 0, 0) = 0.5 \times (1 - 0.4)$.] Compute the probability that a customer buys all three books under $\tilde{P}$:

$$\tilde{P}_{ABC}(1, 1, 1)$$

c. Al Bayes, who fills out the orders, notices that in fact women buy according to model $\bar{P}$ while non-female customers buy according to model $\tilde{P}$.

The next customer, Robin Hood, has ordered books A, B but not C. Using $\bar{P}$ and $\tilde{P}$ from above, decide by the likelihood ratio if Robin Hood is a woman.

e. Al also knows from experience that the prior probability that a customer is a woman is 2/3.

Determine the posterior probability that Robin Hood is a woman, i.e

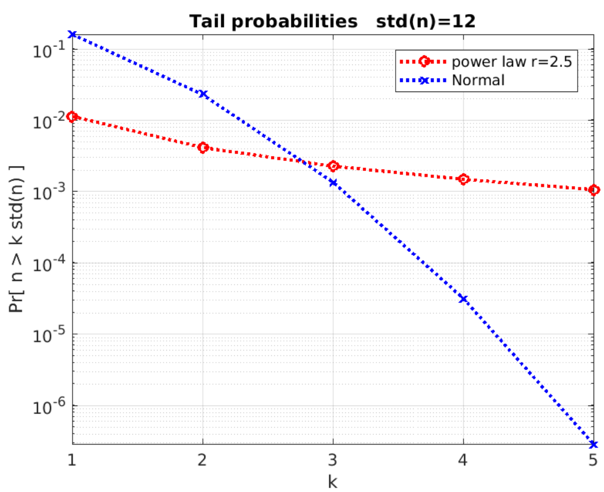
$$P(\text{woman}|A = 1, B = 1, C = 0)$$

f. Determine the the posterior probability that Robin Hood is a woman, if Al, being forgetful, doesn't recall whether Robin ordered book C or not, but he is sure that (s)he ordered A and B.

$$P(\text{woman}|A = 1, B = 1)$$

3 – Heavy tails vs light tails

a topic we did not discuss this year Chapter 8.8 describes a *heavy tailed* distribution, the Power Law distribution. By contrast, distributions like the Normal or Exponential are called *light-tailed.* The term refers to the rate of decay of the "tails" of the distributions towards infinity. Light-tailed distributions decay exponentially, or faster; heavy-tailed distributions decay polinomially, i.e. like x^{-r}, where $r > 0$. The figure (with $r = 2.5$ below illustrates why the difference is important.



Call a *a rare outcome* an x which is k standard deviations away from the mean, or more. The figure shows on a logarithmic scale the probability of rare outcomes for the Normal vs. a Power Law distribution. You can see that, for $k = 3$, the rare outcomes have about the same probability, just over 10^{-3}. Think of k as the cost of a bug, size of a stock market crash, magnitude of an earthquake[1], etc. Outcomes with $k \geq 5$ are 1 in about 10 million under the normal distribution, but have a probability barely half of the outocomes with $k = 3$ under power law. In other words, under a normal distribution, really extreme outcomes are very rare, while under Power Law, once an outcome is "large" it can just as well be "very large".

[1]Earthquakes, fortunately, have light-tailed distributions.