Prof. Dr. Jens Honore Walther
Dr. Julija Zavadlav
ETH Zentrum, CLT
CH-8092 Zürich

Solution Set 13

Issued: 03.05.2019

Question 1: Gaussian Distribution

a) The mean of X is given by,

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx \\ &= \int_{-\infty}^{+\infty} (x-\mu) f_X(x) dx + \mu \int_{-\infty}^{+\infty} f_X(x) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2\sigma^2}} dy + \mu \\ &= \mu,\end{aligned}$$

where we have used that $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ because f_X is a density and that $\int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2\sigma^2}} dy = 0$ due to symmetry. The variance of X is given by,

$$\mathbb{E}\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x) dx.$$

By setting $y = x - \mu$ and $a = \frac{1}{2\sigma^2}$, we obtain

$$\begin{aligned}\mathbb{E}\left[(X-\mu)^2\right] &= \sqrt{\frac{a}{\pi}} \int_{-\infty}^{+\infty} y^2 e^{-ay^2} dy \\ &= -\sqrt{\frac{a}{\pi}} \frac{\partial}{\partial a} \int_{-\infty}^{+\infty} e^{-ay^2} dy \\ &= -\sqrt{\frac{a}{\pi}} \frac{\partial}{\partial a} \sqrt{\frac{\pi}{a}} \\ &= \frac{1}{2a} = \sigma^2.\end{aligned}$$

b) We apply the inversion method to the density of the standard normal distribution,

$$p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

The cumulative distribution can be expanded as follows

$$\begin{aligned} F_X(x) &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t, \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t, \\ &= \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int\limits_{0}^{x/\sqrt{2}} \exp\left(-t^2\right) \mathrm{d}t, \\ &= \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right). \end{aligned}$$

Substituting $x = \sqrt{2}\,\mathrm{erf}^{-1}(2u-1)$,

$$u = F_X(x),$$

which gives the desired result. In practice, the inverse of the error function is not available or not cheap to compute. The Box-Muller algorithm is much cheaper to perform.

Question 2: Bayesian Inference

a) Given the chain rule we can directly see that

$$P(X,Y) = P(X|Y)P(Y) \tag{1}$$

or another valid factorization is

$$P(X,Y) = P(Y|X)P(X) \tag{2}$$

Combining the two yields

$$P(X|Y)P(Y) = P(Y|X)P(X) \tag{3}$$

This can readily be rewritten as Bayes rule.

b) First we define the random variable Y, which takes the value 0 in case the event is *selfdriving car*, and 1 otherwise. Then we define X, taking the value 0 in case a pedestrian has been hit, 1 otherwise. The probabilities required for Bayes rule are:

- $P(Y=0) = 0.1$
- $P(X=0|Y=0) = 0.00005$
- $P(X=0) = \sum_x P(X=0|Y=x)P(Y=x) \approx 0.00005 \cdot 0.1 + 0.00007 \cdot 0.9$

Hence, by simply completing Bayes rule, we have:

$$P(Y=0|X=0) \approx \frac{0.00005 \cdot 0.1}{0.000065} \approx 0.077$$

c) In order to marginalize $P(X_{1:N})$ one has to compute

$$P(X_k) = \int_\Omega P(X_1 = x_1, \ldots, X_k, \ldots, X_N = x_N) d\Omega$$

with $\Omega = \Omega_1 \times \cdots \times \Omega_{k-1} \times \Omega_{k+1} \times \ldots \Omega_N$. If the number of parameters is high (ie. in big data or machine learning), Monte Carlo integration is a common approach to solve the integral.

Question 3: Uncertainty Quantification

a) We want to compute the maximum of $P(q|D)$. In order to simplify the computation we use the fact that the logarithm is a monotonous function and therefore we obtain the same result when optimizing $\log[P(q|D)]$. Another way to see this follows from the chain rule $(\ln(f))' = \frac{f'}{f}$, thus they have the same optimums. Computing the $\log$ of our likelihood gives

$$\log[P(q|D)] = \text{const} + k\log(q) + (n-k)\log(1-q) \tag{4}$$

In order to determine the location of the optima we take the derivative

$$\frac{\mathrm{d}\log[P(q|D)]}{\mathrm{d}q} = \frac{k}{q} - \frac{n-k}{1-q} \tag{5}$$

Setting this to zero allows us to compute the value we where looking for

$$q^* = \frac{k}{n} \tag{6}$$

which is what you expect, namely that it is just the number of heads seen divided by the number of throws. For the given sequence we find $q^* = 0.7$, thus we expect that our coin is unfair.

b) From the lecture we know that we can estimate the standard deviation of an arbitrary probability distribution via the second derivative evaluated at the most probable value. As before we may consider the logarithm to simplify computation

$$\sigma^2 = -\left(\left.\frac{\mathrm{d}^2\log[P(q|D)]}{\mathrm{d}H^2}\right|_{q^*}\right)^{-1} \tag{7}$$

Let us calculate the second derivative

$$\left.\frac{\mathrm{d}^2\log[P(q|D)]}{\mathrm{d}H^2}\right|_{q^*} = -\frac{k}{q^{*2}} - \frac{(n-k)}{(1-q^*)^2} \tag{8}$$

A little bit of algebra and taking the reciprocal value gives

$$\sigma^2 = \frac{q^{*2}(1-q^*)^2}{nq^{*2} - 2q^*k + k} = \frac{q^*(1-q^*)}{n} \tag{9}$$

Where for the second expression we replaced $k = q^*n$. Plugging in $q^* = 0.7$, $k = 7$ and $n = 10$ gives $\sigma = 0.14$. Indeed this is very useful, since by adding more and more measurements the standard deviation goes down (observe the $1/\sqrt{n}$ dependence). Thus this allows us to define an effective threshold after which we are convenient of our estimated parameter.

c) We can use the Monte Carlo methods introduced in class to generate samples from an arbitrary distribution (for example from the binomial distribution used above).