Technische Universität Berlin
Fakultät IV – Elektrotechnik und Informatik

Probabilistic and Bayesian Modelling
in Machine Learning and Artificial Intelligence

Manfred Opper, Théo Galy–Fajou

Summer Term 2018

Problem Sheet 1

Solutions

Problem 1 – Random experiments

A dice is thrown repeatedly until it shows a 6. Let T be the number of throws for this to happen. Obviously, T is a random variable. Compute the expectation value $\mathrm{E}[T]$ and the variance $\mathrm{Var}(T)$ of T.

The probability for t throws is given by the geometric distribution

$$P(T = t) = (1 - q)q^{t-1}$$

with parameter $q = 5/6$. The expectation value of T can be calculated using its definition:

$$\mathrm{E}[T] = \sum_{t=1}^{\infty} t\, P(T = t) = \sum_{t=1}^{\infty}(1 - q)\, t\, q^{t-1} = \sum_{t=1}^{\infty}(1 - q)\frac{d}{dq}\, q^t$$

As the geometric series converges absolutely, we can exchange summation and derivation:

$$\mathrm{E}[T] = (1 - q)\frac{d}{dq}\sum_{t=0}^{\infty} q^t = (1 - q)\frac{d}{dq}\frac{1}{1 - q} = (1 - q)\frac{1}{(1 - q)^2} = \frac{1}{1 - q}$$

In order to obtain the variance we need the expectation value of T^2, too:

$$\mathrm{E}[T^2] = \sum_{t=1}^{\infty} t^2\, P(T = t) = \sum_{t=1}^{\infty}(1 - q)\, t^2\, q^{t-1}$$

Here $t^2\, q^{t-1}$ is very similar to the second derivative of q^{t+1}:

$$\mathrm{E}[T^2] = \sum_{t=1}^{\infty}(1 - q)\, t(t + 1)q^{t-1} - \sum_{t=1}^{\infty}(1 - q)tq^{t-1} = -\mathrm{E}[T] + \sum_{t=1}^{\infty}(1 - q)\frac{d^2}{dq^2}\, q^{t+1}$$

Further simplifications

$$\mathrm{E}[T^2] = -\frac{1}{1-q} + (1-q)\frac{d^2}{dq^2}\sum_{t=0}^{\infty} q^t = -\frac{1}{1-q} + (1-q)\frac{d^2}{dq^2}\frac{1}{1-q}$$

lead to

$$\mathrm{E}[T^2] = -\frac{1}{1-q} + (1-q)\frac{2}{(1-q)^3} = \frac{1+q}{(1-q)^2}$$

so that the variance of T is given by

$$\mathrm{Var}(T) = \mathrm{E}[T^2] - \mathrm{E}[T]^2 = \frac{1+q}{(1-q)^2} - \frac{1}{(1-q)^2} = \frac{q}{(1-q)^2}$$

By substituting $q = 5/6$ we finally find $\mathrm{E}[T] = 6$ and $\mathrm{Var}(T) = 30$.

Problem 2 – Addition of variances

Let X and Y be independent random variables. Show that

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y),$$

where the variance is defined as

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

Hint: Use the fact that for independent U and V, $\mathrm{E}[UV] = \mathrm{E}[U]\mathrm{E}[V]$.

$$\begin{aligned}
& \mathrm{Var}(X+Y) \\
= \; & \mathrm{E}[(X+Y-E[X+Y])^2] \\
= \; & \mathrm{E}[X^2 + Y^2 + \mathrm{E}[X+Y]^2 + 2XY - 2X\mathrm{E}[X+Y] - 2Y\mathrm{E}[X+Y]] \\
= \; & \mathrm{E}[X^2] + \mathrm{E}[Y^2] + \mathrm{E}[X+Y]^2 + 2\mathrm{E}[XY] \\
& -2\mathrm{E}[X]\mathrm{E}[X+Y] - 2\mathrm{E}[Y]\mathrm{E}[X+Y] \\
= \; & \mathrm{E}[X^2] + \mathrm{E}[Y^2] + \mathrm{E}[X]^2 + 2\mathrm{E}[X]\mathrm{E}[Y] + \mathrm{E}[Y]^2 + 2\mathrm{E}[X]\mathrm{E}[Y] \\
& -2\mathrm{E}[X]^2 - 2\mathrm{E}[X]\mathrm{E}[Y] - 2\mathrm{E}[X]\mathrm{E}[Y] - 2\mathrm{E}[Y]^2 \\
= \; & \mathrm{E}[X^2] - \mathrm{E}[X]^2 + \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 \\
= \; & \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

The last step uses the computational formula for the variance:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[(X - \mathrm{E}[X])^2] \\
&= \mathrm{E}[X^2 - 2X\mathrm{E}[X] + \mathrm{E}[X]^2] \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}[X]^2 \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2
\end{aligned}$$

Problem 3 – Transformation of probability densities

Let X be uniformly distributed in $(0, 1)$:

$$p(x) = \begin{cases} 1 & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

A second random variable Y is defined as

$$Y = \tan\left(\pi(X - 1/2)\right).$$

What is the probability density $q(y)$ of Y?

- Inverse function:

$$\begin{aligned} y = \tan(\pi(x - 1/2)) \quad &\Longleftrightarrow \quad \arctan y = \pi(x - 1/2) \\ &\Longleftrightarrow \quad x = \frac{1}{\pi} \arctan y + \frac{1}{2} \end{aligned}$$

- Transformation of probability densities:

$$q(y) = p(x) \cdot \frac{dx}{dy} = p(x) \cdot \frac{1}{\pi} \frac{1}{1 + y^2} = \frac{1}{\pi} \frac{1}{1 + y^2}$$

- This transformation together with a (pseudo-)random number generator can be used to generate (pseudo-)random numbers with a standard Cauchy distribution.

Problem 4 – Gaussian inference

Suppose we have two random variables V_1 and V_2 which are **jointly Gaussian** distributed with zero means $E[V_1] = E[V_2] = 0$ and variances $E[V_1^2] = 16.6$ and $E[V_2^2] = 6.8$. The covariance is $E[V_1\, V_2] = 6.4$.

Assume that we observe a noisy estimate

$$Y = V_2 + \nu$$

of V_2 where ν is a Gaussian noise variable independent of V_1 and of V_2 with $E[\nu] = 0$ and $E[\nu^2] = 1$.

(a) Calculate the conditional (posterior) densities $p(V_1|Y)$ and $p(V_2|Y)$.

(b) What are the posterior mean predictions of V_1 and V_2 for an observation $Y = 1$ and what are the posterior uncertainties of these predictions.

The following formula could be helpful: The inverse of the matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

is given by

$$\mathbf{A}^{-1} = \frac{1}{\det\mathbf{A}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

(a) We obtain the conditional densities $p(V|Y)$ from the joint densities $p(V,Y)$. (Here V can be either V_1 or V_2) !

$$p(V,Y) = \frac{1}{2\pi\sqrt{\det(\mathbf{S})}} \exp\left\{-\frac{1}{2}(V,\ Y)^\top \mathbf{S}^{-1}(V,\ Y)\right\}$$

Note $(V,\ Y)$ is a two dimensional vector and the covariance matrix is given by

$$\mathbf{S} = \begin{pmatrix} E[V^2] & E[VY] \\ E[VY] & E[Y^2] \end{pmatrix}$$

The expectations are

$$\begin{aligned} E[V_1Y] &= E[V_1V_2] \\ E[V_2Y] &= E[V_2^2] \\ E[Y^2] &= E[V_2^2] + E[\nu^2] \end{aligned}$$

We set

$$\mathbf{S}^{-1} = \begin{pmatrix} (\mathbf{S}^{-1})_{vv} & (\mathbf{S}^{-1})_{vy} \\ (\mathbf{S}^{-1})_{vy} & (\mathbf{S}^{-1})_{yy} \end{pmatrix}$$

Then, from the joint density, we can write the conditional density as

$$p(V|Y) \propto \exp\left(-\frac{V^2}{2}(\mathbf{S}^{-1})_{vv} - V(\mathbf{S}^{-1})_{vy}Y\right)$$

(b) This can be written in the standard notation as

$$p(V|Y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(V-\mu)^2}{2\sigma^2}}$$

where

$$\mu = E[V|Y] = -\frac{(\mathbf{S}^{-1})_{vy}Y}{(\mathbf{S}^{-1})_{vv}}$$

$$\sigma^2 = \mathrm{VAR}[V|Y] = \frac{1}{(\mathbf{S}^{-1})_{vv}}$$

are the conditional mean and variance. We can use $E[V|Y]$ for prediction. $\mathrm{VAR}[V|Y]$ would give us a measure for the error of such a prediction.

(c) Numerical example:

- For $p(V_1|Y)$ we have

$$\mathbf{S} = \begin{pmatrix} 16.6 & 6.4 \\ 6.4 & 7.8 \end{pmatrix}$$

  and

$$\mathbf{S}^{-1} = \begin{pmatrix} 0.0881 & -0.0723 \\ -0.0723 & 0.1875 \end{pmatrix}$$

  Hence $E[V_1|Y] = 0.8207$ and $\mathrm{VAR}[V_1|Y] = 11.3507$.

- For $p(V_2|Y)$ we have

$$\mathbf{S} = \begin{pmatrix} 6.8 & 6.8 \\ 6.8 & 7.8 \end{pmatrix}$$

  and

$$\mathbf{S}^{-1} = \begin{pmatrix} 1.1471 & -1.0000 \\ -1.0000 & 1.0000 \end{pmatrix}$$

  Hence $E[V_2|Y] = 0.8718$ and $\mathrm{VAR}[V_2|Y] = 0.8718$.