Information Theory & Statistical Physics

Lecture: Prof. Dr. Johannes Berg
Exercises: Stephan Kleinbölting **Sheet 0**

Due date: – *To be discussed on: 26.04.17*

1 Probabilities

Let X be a random variable that takes values in a set of outcomes $\mathcal{H}$. One often has $\mathcal{H} = \mathbb{N}, \mathbb{R}, \mathbb{R}^+, [a, b], \ldots$

Talking about the probability $\mathrm{P}(X = x)$ that X takes a specific value $x \in \mathcal{H}$ works well if $\mathcal{H}$ is a countable set. On the other hand, if $\mathcal{H}$ is not countable, e.g $\mathcal{H} = \mathbb{R}$, this approach fails because any single point necessarily carries probability 0 (Think about this!) For this and – more importantly – practical reasons, one would like to consider not only single outcomes, but whole subsets of outcomes, so called *events*. Think of rolling a six-faced die and obtaining an even/odd number. Or rolling a thousand dice and considering the event "the sum of eyes is larger than 3700".

Let us write $\mathcal{F}$ for the set of all events. If such a system of subsets is suitable for modeling real-world events it should at least be compatible with the operations of propositional logic (AND, OR, NOT), i.e we would like to talk about *(A AND B)*, or *(A OR B AND NOT C)* happening. We therefore demand that the following axioms hold

(1) **The event "anything" can happen is included:** $\mathcal{H} \in \mathcal{F}$

(2) **Negation:** $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$. If A can happen, so must $A^c \equiv \mathcal{H} \setminus A$.

(3) **Logical OR:** For I a (at most) countable index set and $A_i \in \mathcal{F}$: $\bigcup_{i \in I} A_i \in \mathcal{F}$

N.B.: A system of sets obeying these axioms is called a *σ-algebra.*

(a) Show that (1)–(3) imply

i. $\emptyset \in \mathcal{F}$

ii. $A \cap B \in F$

iii. $A \setminus B \in F$

iv. *Bonus:* $\bigcap_{i \in I} A_i \in \mathcal{F}$

Finally we assign probabilities $\mathrm{P} : \mathcal{F} \to [0, 1]$ to the events. For consistency we require that

(1) **Something will happen surely:** $\mathrm{P}(\mathcal{H}) = 1$

(2) **Probabilities of mutually exclusive events add up:** If A_i are disjoint, then $\mathrm{P}(\bigcup_{i \in I} A_i) = \sum_{i \in I} \mathrm{P}(A_i)$.

(b) Show that

i. $\mathrm{P}(\emptyset) = 0$

ii. $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$

iii. $\mathrm{P}(A \setminus B) = \mathrm{P}(A) - \mathrm{P}(A \cap B)$

iv. For not necessarily disjoint $A, B \in \mathcal{F}$: $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

Hint: Start with $\mathcal{H}$ and decompose it into suitable disjoint subsets. It might be helpful to draw a picture.

Solution

(a) i. $\mathcal{H} \in \mathcal{F}$, and by axiom ii. $\mathcal{H}^c = \emptyset \in \mathcal{F}$.

ii. Write $A \cap B$ as the complement of everything that is *not* in $A \cap B$:

$$A \cap B = (A^c \cup B^c)^c \in \mathcal{F}$$

This is one of De Morgan's laws. They are valid for arbitrary intersections, which also solves iv.

iii.

$$A \setminus B = \{x \in A | x \notin B\} = A \cap B^c \in \mathcal{F}$$

(b) i.

$$1 = P(\mathcal{H}) = P(\mathcal{H} \cup \emptyset) = P(\mathcal{H}) + P(\emptyset) = 1 + P(\emptyset)$$

The second equality is due to $\mathcal{H}$ and $\emptyset$ being disjoint.

ii.

$$1 = P(\mathcal{H}) = P(A^c \cup A) = P(A^c) + P(A)$$

iii.

$$\begin{aligned} 1 &= P((A \setminus B) \cup (A \setminus B))^c = P(A \setminus B) + P((A \setminus B)^c) \\ &= P(A \setminus B) + P(A^c \cup A \cap B) = P(A \setminus B) + \underbrace{P(A^c)}_{1-P(A)} + P(A \cap B) \end{aligned}$$

iv.

$$\begin{aligned} P(A \cup B) &= P(A \setminus B \cup B \setminus A \cup A \cap B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= 2P(A \cup B) - P(A) - P(B) + P(A \cap B) \end{aligned}$$

Note: A tempting choice for $\mathcal{F}$ seems to be the powerset over $\mathcal{H}$, which is the set of all subsets, and trivially fulfills the axioms. This works well in many cases, but fails when trying to define probabilities if $\mathcal{H}$ is not countable.

2 Probability distributions

(Real) random variables come in two flavors: discrete and continuous. Discrete RVs take values in a countable set, e.g the outcome of rolling a die, or the number of radioactive decays during some time interval. Continuous RVs can take arbitrary values in (some suitable subset of) the real numbers.

Let X, Y be a discrete and a continuous random variable that take values in $\mathcal{A}_X$and $\mathcal{A}_Y \subseteq \mathbb{R}$, respectively. For a discrete variable it makes sense to talk about the probability that X takes a particular value $x \in \mathcal{A}_X$. We call this function $p : \mathcal{A}_X \to [0, 1]$

$$p(x) := \mathrm{P}(X = x) \equiv \mathrm{P}(X \in \{x\})$$

the **Probability Mass Function (PMF)**.

Let $N \in \mathbb{N}, 0 \leq f \leq 1$ and X distributed according to

$$p(k) = \binom{N}{k} f^k (1-f)^{N-k} \tag{1}$$

which is the PMF of a Binomial distribution.

(a) Calculate the expectation value $\langle X \rangle = \sum_{k=0}^{\infty} kp(k)$ and variance $\mathrm{Var}(X) = \langle (X - \langle X \rangle)^2 \rangle$ of X. *Hint:* First show that $\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2$ for any random variable X.

The probability for a continuous RV to take on one specific value is zero. One therefore considers intervals $(a, b]$ and evaluates the probability $\mathrm{P}(X \in (a, b])$. In particular one commonly considers the half-open intervals $(-\infty, y]$, and thus defines the **Cumulative Distribution Function (CDF)** of Y $F_Y : \mathbb{R} \to [0, 1]$

$$F_Y(y) = \mathrm{P}(Y \in (-\infty, y]) \equiv \mathrm{P}(Y \leq y)$$

(b) Show that for $a \leq b$

$$\mathrm{P}(a < Y \leq b) = F_Y(b) - F_Y(a) \tag{2}$$

If F_Y is nicely behaved (absolutely continuous), it possesses a **Probability Density Function (PDF)**

$$F_Y(y) = \int_{-\infty}^{y} p(x) dx \tag{3}$$

Obviously

$$p(x) = \frac{dF_Y}{dy}(x) \tag{4}$$

wherever F_Y is differentiable.

Let Y be exponentially distributed on $\mathbb{R}^+$ with PDF

$$p(y) = \lambda e^{-\lambda y}, \quad \lambda > 0 \tag{5}$$

(c) Calculate the expection value and variance of Y.

(d) What is the probability for $1 < Y < 2$?

(e) Let

$$F(y) = \begin{cases} 0 & y \leq 0 \\ \sqrt{y} & 0 < y < 1 \\ 1 & y \geq 1 \end{cases} \tag{6}$$

What is the corresponding PDF?

Solution

(a)

(b)

$$\begin{aligned} P(a \leq Y \leq b) &= P((-\infty, b] \setminus (-\infty, a]) = P((-\infty, b]) - P((-\infty, b] \cap (-\infty, a]) \\ &= F_Y(b) - P((-\infty, a]) = F_Y(b) - F_Y(a) \end{aligned}$$

(c)

$$\begin{aligned} \langle X \rangle &= \int_0^{\infty} x \lambda e^{-\lambda x} dx = -\int_0^{\infty} x \frac{d}{dx} e^{-\lambda x} dx \\ &= -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda} \end{aligned}$$

Calculating $\langle X^2 \rangle$ works in the same manner.

(d) By integrating the PDF on finds the CDF

$$F_Y(y) = 1 - e^{-\lambda y}$$

and hence

$$P(1 < Y < 2) = e^{-\lambda} - e^{-2\lambda} > 0$$

(e) The CDF is everywhere but at $\{0, 1\}$ differentiable. We thus find the PDF

$$p_Y(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2\sqrt{x}} & 0 < x < 1 \\ 0 & x > 1 \end{cases}$$

It is not defined at $\{0, 1\}$, but that is not a problem, since the (Lebesgue) integral is oblivious to isolated points.

3 Joint distributions and covariance

(a) Consider an urn filled with N balls, half of which are black, the other half white. We draw two balls without replacement. Let's denote the outcomes $X, Y \in \{b, w\}$. As a twist we have the following rule: If the first draw turns up a white ball $X = w$, discard all remaining white balls. The next draw will then be a black ball with certainty.

i. Write down the conditional distribution $p(y|x)$.

ii. Calculate the joint distribution $p(x, y)$ and the marginal $p(y)$. Make a table! What is the probability to get a white ball in the second draw for large N?

Let $\mathrm{Cov}(X, Y) = \langle (X - \langle X\rangle)(Y - \langle Y\rangle)\rangle$ denote the *covariance* of X and Y.

(b) Show that for random variables X, Y, Z and $a \in \mathbb{R}$

i. $\langle aY + X\rangle = a\langle Y\rangle + \langle X\rangle$

ii. $\mathrm{Cov}(aX + Y, Z) = a\mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$

It follows from symmetry that the covariance is also linear in its second argument.

(c) Show that for any independent X, Y, the covariance vanishes. Note that the reverse does not hold in general! You might want to show first that for X, Y independent $\langle XY\rangle = \langle X\rangle\langle Y\rangle$.

Solution

(a) First note that $P(X = b) = P(X = w) = 1/2$.

$P(Y\|X)$ X Y	b	w	$P(Y, X)$ X Y	b	w
b	$\frac{N/2-1}{N-1}$	$\frac{N/2}{N-1}$	b	$\frac{1}{4}\frac{N-2}{N-1}$	$\frac{1}{4}\frac{N}{N-1}$
w	1	0	w	$\frac{1}{2}$	0

$P(Y = w) = \frac{1}{4}\frac{N}{N-1} \to \frac{1}{4}$ is found by summing over the second column of $P(X, Y)$.

(b) i.

$$\begin{aligned}\langle aX + Y\rangle &= \sum_x \sum_y ax + yp(x, y) \\ &= a\sum_x x \sum_y p(x, y) + \sum_y y \sum_x p(x, y) = a\sum_x xp(x) + \sum_y yp(y)\end{aligned}$$

ii.

$$\begin{aligned}\mathrm{Cov}(aX+Y,Z) &= \langle (aX+Y-\langle aX+Y\rangle)(Z-\langle Z\rangle)\rangle \\ &= \langle a(X-\langle X\rangle)(Z-\langle Z\rangle)+(Y-\langle Y\rangle)(Z-\langle Z\rangle)\rangle \\ &= a\mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z)\end{aligned}$$

As a corollary one finds $\mathrm{Var}(aX)=\mathrm{Cov}(aX,aX)=a^2\mathrm{Var}(X)$.

(c)

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \langle XY\rangle - \langle X\rangle\langle Y\rangle \\ &= \sum_{x,y} xyp_{XY}(x,y) - \langle X\rangle\langle Y\rangle \\ &= \sum_x xp_X(x)\sum_y yp_Y(y) - \langle X\rangle\langle Y\rangle = 0\end{aligned}$$

The reverse is not true in general: Take X uniformly distributed on $[-1,1]$ (i.e X has PDF $p_X(x)=1/2$ on $[-1,1]$), and $Y=X^2$. X and Y are clearly dependent, since X fixes Y completely. Nevertheless $\mathrm{Cov}(X,X^2)=0$ as is easily demonstrated.

4 Multivariate normal distribution

You are likely familiar with the univariate normal distribution

$$p(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma}} \tag{7}$$

It has expectation value μ and variance σ. One writes $X\sim\mathcal{N}(\mu,\sigma)$ for a random variable distributed in this way.

Let $X_1,X_2\sim\mathcal{N}(0,1)$ be independent and identically (iid.) standard-normal distributed, and consider the joint PDF of $\mathbf{X}=(X_1,X_2)$, $p(\mathbf{x})=p(x_1,x_2)=p(x_1)p(x_2)$.

Next we correlate X_1 and X_2 by linearly combining them into a new random vector $\mathbf{Y}$

$$\mathbf{Y}=\mathbf{B}\cdot\mathbf{X} \tag{8}$$

$\mathbf{B}$ is a a real invertible matrix.

(a) Use the transformation law for densities (which we use here without proof)

$$p_{\mathbf{Y}}(\mathbf{y})=p_{\mathbf{X}}(\mathbf{x}(\mathbf{y}))\left|\frac{d\mathbf{x}}{d\mathbf{y}}(\mathbf{y})\right| \tag{9}$$

to show that $\mathbf{Y}$ is distributed according to

$$p(\mathbf{y})=\frac{1}{2\pi\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right) \tag{10}$$

where Σ is a real symmetric and positive 2x2 matrix.

(b) Calculate the covariance $\mathrm{Cov}(Y_1,Y_2)$.
Hint: You do not need to explicitely calculate the integral.

(c) Show that Y_1,Y_2 are independent iff. Σ is diagonal.

(d) *Bonus:* Show that the marginal distributions $p(x)=\int_{\mathbb{R}}p(x,y)dy$ are again normally distributed.
Hint: Explicitely calculate Σ^{-1}.

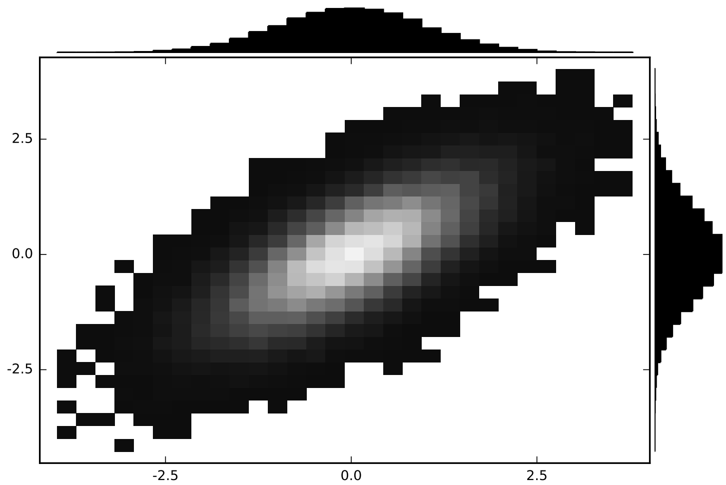

Figure 1: Histogram and marginals from a bivariate Gaussian with $\Sigma = \begin{pmatrix} 1.0 & 0.7 \\ 0.7 & 1.0 \end{pmatrix}$

Solution

(a) Obviously $\mathbf{x} = \mathbf{B^{-1}x}$. The Jacobian of the transformation is simply $\mathbf{B^{-1}}$. Defining $\Sigma = \mathbf{BB^T}$, and noting that $\det \Sigma = \det \mathbf{BB^T} = (\det \mathbf{B})^2 = (\det \mathbf{B^{-1}})^{-2}$, the result follows. Σ is by construction real symmetric and invertible.

(b) Let's do the calculation for any dimension of $\mathbf{B}$.

$$\begin{aligned} \mathrm{Cov}(Y_i, Y_j) &= \mathrm{Cov}(\sum_k B_{ik} X_k, \sum_l B_{jl} X_l) \\ &= \sum_{k,l} B_{ik} B^T_{lj} \mathrm{Cov}(X_k, X_l) \\ &= (BB^T)_{ij} = \Sigma_{ij} \end{aligned}$$

The second to last equality follows because due to independence $\mathrm{Cov}(X_k, X_l) = \delta_{kl}$. Note that this result does not rely on any specifics of the distribution, and is thus valid in general.

(c) If Σ is diagonal the joint distribution immediately factorizes. On the other hand, independence implies vanishing covariance and due to b) also the vanishing of the off-diagonal entries of Σ.

(d) Invert Σ^{-1} and expand the exponential. You are left with a Gaussian integral, that can be performed to yield another Gaussian.