Avd. Matematisk statistik

KTH Matematik

EXAM IN SF2955 COMPUTER INTENSIVE METHODS
WEDNESDAY 25th of May 2011 14.00 – 19.00 in V34.

Examiner: Timo Koski, tel. 790 71 34. Email: `tjtkoski@kth.se`

Allowed aids: Formel- och tabellsamling i Matematisk statistik. Formulas in Bayesian Inference. Pocket calculator.

There are six (6) assignments (uppgift). Notation should be explained and defined. Arguments must be detailed enough to make it possible to follow. Numerical answers should be given with the precision of at least 2 significant digits. Solutions written in Swedish are welcome.

Results will be ready before June the 1st.

Good luck!

- -

Uppgift 1

$\mathbf{X} = (X_1, X_2, X_3)$ is a vector of three binary random variables with the probability mass function on any $\mathbf{x} = (x_1, x_2, x_3)$ conditionally on $\Theta_1 = \theta_1, , \Theta_2 = \theta_2, \Theta_3 = \theta_3$ given by

$$p_{X_1,X_2,X_3|\Theta_1,\Theta_2,\Theta_3}(\mathbf{x} \mid \theta_1, \theta_2, \theta_3) = \prod_{i=1}^{3} p_{X_i}(x_i \mid \theta_i),$$

where for $x_i \in \{0, 1\}$

$$p_{X_i}(x_i \mid \theta_i) = \theta_i^{x_i}(1 - \theta_i)^{1-x_i},$$

so that $p_{X_i}(x_i \mid \theta_i) = \theta_i$ and $p_{X_i}(0 \mid \theta_i) = 1 - \theta_i$. We have four independent samples of $\mathbf{X}$ denoted by $\mathbf{x}^{(l)}$ for $l = 1, 2, 3, 4$, and given by

	x_1	x_2	x_3
$\mathbf{x}^{(1)}$	0	0	1
$\mathbf{x}^{(2)}$	0	1	0
$\mathbf{x}^{(3)}$	1	1	1
$\mathbf{x}^{(4)}$	0	1	1

The Bayesian Ansatz is that each Θ_i has the prior density $\mathcal{B}\text{eta}(1/2, 1/2)$, i.e.

$$\pi_{\Theta_i}(\theta_i) = \begin{cases} \frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)} \theta_i^{-1/2}(1 - \theta_i)^{-1/2} & 0 < \theta_i < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

and that Θ_1, Θ_2 and Θ_3 are independent random variables.

Compute the posteriori density for $\Theta_1, \Theta_2, \Theta_3$ given $\mathbf{x}^{(l)}$ for $l = 1, 2, 3, 4,$, i.e.,

$$f_{\Theta_1,\Theta_2,\Theta_3|\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)},\mathbf{x}^{(4)}}(\theta_1, \theta_2, \theta_3).$$

(10 p)

Uppgift 2

We want to estimate the parameter $\theta = P(X > 1)$ and have for this purpose access to 20 independent observations/samples of X:

0.07, 0.07, 0.17, 0.19, 0.38, 0.39, 0.39, 0.40, 0.53, 0.64,
0.66, 0.87, 0.96, 0.96, 1.52, 2.71, 2.93, 3.06, 3.36, 4.87.

a) Find the plug-in-estimate of θ numerically. What is the distribution of this estimator (expressed in θ) ? (5 p)

b) We bootstrap this plug-in-estimate. Give the distribution of this bootstrap estimate. (5 p)

Uppgift 3

We consider the Gibbs' sampler for simulating a probability density $\pi_{X_1,X_2}(x_1, x_2)$ for a bivariate random variable X_1, X_2. Suppose that the Markov chain is currently in the state (x_1, x_2).
i) We update first x_1 by drawing y_1 from the proposal probability density for proposing (y_1, y_2) defined by

$$q\left((x_1, x_2), (y_1, y_2)\right) = \begin{cases} \pi_{X_1|X_2=x_2}(y_1) & \text{if } y_2 = x_2 \\ 0 & \text{if } y_2 \neq x_2. \end{cases}$$

ii) The acceptance probability for (y_1, y_2) is

$$\alpha\left((x_1, x_2), (y_1, y_2)\right) = \min\left(1, \frac{\pi_{X_1,X_2}(y_1, y_2)\, q\left((y_1, y_2), (x_1, x_2)\right)}{\pi_{X_1,X_2}(x_1, x_2)\, q\left((x_1, x_2), (y_1, y_2)\right)}\right).$$

a) Show that the acceptance probability above is always equal to = 1. (6 p)

b) What is the next step of the Gibbs' sampler ? Give the proposal density for this step. (4 p)

Uppgift 4

Let $x_1, \ldots, x_n$ be I.I.D. samples of a random variable X. The parameter of interest is $\theta = E[X]$. We take B independent bootstrap samples

$$x_1^*(b), \ldots, x_n^*(b), \quad b = 1, 2, \ldots, B$$

of $x_1, \ldots, x_n$. We calculate for each bootstrap replica the arithmetic mean

$$\overline{x^*}(b) = \frac{1}{n}\sum_{i=1}^{n} x_i^*(b),$$

and then set

$$\overline{x^*} = \frac{1}{B}\sum_{b=1}^{B} \overline{x^*}(b).$$

We are interested in studying the bootstrap hypothesis and estimate the bias of $\overline{x^*}$ by

$$\text{BIAS} = \overline{x^*} - \overline{x},$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

a) Find

$$E_{\widehat{F}_n}\left[\text{BIAS}\right],$$

where $\widehat{F}_n$ is the empirical distribution function corresponding to the n samples $x_1, \ldots, x_n$. Or, $E_{\widehat{F}_n}$ is the expectation over all bootstrap simulations. This is the conditional expectation of BIAS given $x_1, \ldots, x_n$. (2 p)

b) Show that

$$V_{\widehat{F}_n}\left[\text{BIAS}\right] = \frac{\widehat{\sigma}^2}{Bn},$$

where $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$. This is the conditional variance of BIAS given $x_1, \ldots, x_n$. (2 p)

c) The unconditional variance of BIAS is

$$V\left[\text{BIAS}\right],$$

where we are now computing the variance of our bias of bootstrap estimate due to the variability between different samples of X.

Show that

$$V\left[\text{BIAS}\right] = \left(\frac{n-1}{n}\right) \cdot \frac{\sigma^2}{Bn},$$

where $\sigma^2 = V\left[X\right]$.

Aid: For two random variables BIAS and X it holds that

$$V\left[\text{BIAS}\right] = V\left[E(\text{BIAS}|X)\right] + E\left[V(\text{BIAS}|X)\right],$$

where $E(\text{BIAS}|X)$ is the conditional expectation of BIAS given X and $V(\text{BIAS}|X)$ is the conditional variance of BIAS given X . (6 p)

Uppgift 5

Let F be a distribution function on the real line such that

$$F(b) - F(a) = \int_a^b dF(x)$$

for all real numbers $a \leq b$. Set for a fixed pair $a < b$

$$\theta = T(F) = F(b) - F(a).$$

We can write this as

$$T(F) = \int_a^b dF(x) = \int_{-\infty}^{\infty} I_{]a,b]}(x)dF(x),$$

where $I_{]a,b]}(x)$ is the indicator function

$$I_{]a,b]}(x) = \begin{cases} 1 & \text{if } a < x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

a) Show that $T(F)$ is a **linear (statistical) functional**, i.e., if F and G are two distribution functions on the real line, then

$$T(\lambda F + \mu G) = \lambda T(F) + \mu T(G)$$

where λ and μ are two arbitrary real numbers. (2 p)

b) Let δ_t be the distribution function that has all probability mass in the point t. Or,

$$\delta_t(x) = \begin{cases} 0 & x < t \\ 1 & t \leq x. \end{cases}$$

We have for any integrable function $a(x)$ that (*You are not required to prove this*)

$$T(\delta_t) = \int_{-\infty}^{\infty} a\,(x)\,d\delta_t(x) = a(t).$$

The **influence function** $L_T(t;F)$ of T is defined as

$$L_T(t;F) = \lim_{\epsilon \to 0} \frac{T\left((1-\epsilon)F + \epsilon\delta_t\right) - T\left(F\right)}{\epsilon}.$$

Show that

$$L_T(t;F) = I_{]a,b]}(t) - \left(F(b) - F(a)\right).$$

(6 p)

c) The **empirical influence function** $\widehat{L}_T(t)$ of T is defined as $\widehat{L}_T(t) = L_T(t;\widehat{F}_n)$, or,

$$\widehat{L}_T(t) = \lim_{\epsilon \to 0} \frac{T\left((1-\epsilon)\widehat{F}_n + \epsilon\delta_t\right) - T\left(\widehat{F}_n\right)}{\epsilon}$$

where $\widehat{F}_n$ is the empirical distribution function corresponding to n samples $x_1, \ldots, x_n$ from F. Find $\widehat{L}_T(t)$ for $T(F)$. (1 p)

d) What is the delta method in this case ? How is this related to the bootstrap hypothesis ? (1 p)

Uppgift 6

We want to simulate π_1 by a Markov chain $\mathbf{X}^{(1)}$. In order to 'heat up' the Metropolis-Hastings chain $\mathbf{X}^{(1)}$ by a 'faster' Markov chain $\mathbf{X}^{(2)}$ we use a technique known as **Metropolis-Coupled Markov Chain Monte Carlo** a.k.a. MCMCMC a.k.a. MC^3.
The MC^3 works as follows . We are running two parallel Markov chains $\mathbf{X}^{(i)} = \{X_n^{(i)}\}_{n\geq 0}$, $i = 1, 2$, in the same finite discrete state space $\mathcal{S}$ by two independent applications of the Metropolis-Hastings algorithm. The target distribution of $\mathbf{X}^{(1)}$ is π_1 and the proposal distribution matrix is $\mathbf{Q}^{(1)}$. The target distribution of $\mathbf{X}^{(2)}$ is π_2 and the proposal distribution matrix is $\mathbf{Q}^{(2)}$. The proposals and acceptances in the chain $\mathbf{X}^{(1)}$ are done indepedently of the proposals and acceptances in $\mathbf{X}^{(2)}$. Then the steps in the MC^3 algorithm are:

1. Assume $X_{n-1}^{(1)} = u \in \mathcal{S}$ and $X_{n-1}^{(2)} = v \in \mathcal{S}$. Make two independent cycles (proposal, acceptance) of the Metropolis-Hastings algorithm to get $X_n^{(1)} = x_1 \in \mathcal{S}$ and $X_n^{(2)} = x_2 \in \mathcal{S}$.

2. Propose **switch** a.k.a. **swap** of the pair of states x_1, x_2 at random with the probability
$$\alpha\left((x_1, x_2), (x_2, x_1)\right) = \min\left(1, \frac{\pi_1(x_2)\,\pi_2(x_1)}{\pi_1(x_1)\,\pi_2(x_2)}\right).$$
The swap of states is refused with the probability $1 - \alpha\left((x_1, x_2), (x_2, x_1)\right)$.

3. If a swap is accepted, set $X_n^{(1)} \leftarrow x_2$ and $X_n^{(2)} \leftarrow x_1$, set $n \mapsto n-1$, and go to step 1, else go to step 4..

4. If a swap is refused, set $X_n^{(1)} \leftarrow x_1$ and $X_n^{(2)} \leftarrow x_2$ set $n \mapsto n-1$, and go to step 1.

We write this using transition matrices and think that the chain $\mathbf{X}^{(1)}$ is up-dated first (and the second chain waits, so to speak, at $X_{n-1}^{(2)} = v$, although these are parallell steps). For the first up-date the conditional probability of moving from (u, v) to (x_1, x_2) is

$$P^{(1)}\left((u, v), (x_1, x_2)\right) = \begin{cases} P^{(1)}(u, x_1) & \text{if } v = x_2 \\ 0 & \text{if } v \neq x_2. \end{cases}$$

Here $P^{(1)}(u, x_1)$ is the conditional transition probability of the chain $\mathbf{X}^{(1)}$ (and by the theory of Metropolis-Hastings this is $= q^{(1)}(u, x_1)\,\alpha^{(1)}(u, x_1)$, but we do not need this expression here). The corresponding matrix is $\mathbf{P}^{(1)} = \left(P^{(1)}\left((u, v), (x_1, x_2)\right)\right)_{(u,v)\in\mathcal{S},(x_1,x_2)\in\mathcal{S}}$.
Let $\mu\left((x_1, x_2)\right) = \pi_1(x_1)\pi_2(x_2)$ and take this as a row vector μ (taking the pairs (x_1, x_2) as a list with some enumeration).
a) Show that
$$\mu = \mu \mathbf{P}^{(1)}.$$

Aid: Componentwise this means that

$$\mu\left((x_1, x_2)\right) = \sum_{(u,u)} \mu\left((u, v)\right) P^{(1)}\left((u, v), (x_1, x_2)\right).$$

(3 p)

b) One can define the transition probabilities of the chain $\mathbf{X}^{(2)}$ as

$$P^{(2)}\left((u,v),(x_1,x_2)\right) = \begin{cases} P^{(2)}\left(v,x_2\right) & \text{if } u = x_1 \\ 0 & \text{if } u \neq x_1 \end{cases}$$

and then show analogously that $\mu = \mu \mathbf{P}^{(2)}$. You need not do this.
Next we express the swap by means of a transition matrix. We consider $\mathbf{Z} = \{\mathbf{Z}_n = \left(Z_n^{(1)}, Z_n^{(2)}\right)\}_{n \geq 0}$ on the state space $\mathcal{S} \times \mathcal{S}$ such that

$$\mathbf{Z}_n = \left(Z_n^{(1)}, Z_n^{(2)}\right) = \begin{cases} \left(X_n^{(2)}, X_n^{(1)}\right) & \text{with probability } \alpha \\ \left(X_n^{(1)}, X_n^{(2)}\right) & \text{with probability } 1-\alpha. \end{cases}$$

where $\alpha = \alpha\left(\left(X_n^{(1)}, X_n^{(2)}\right), \left(X_n^{(2)}, X_n^{(1)}\right)\right)$.
Find the conditional probability matrix for the swap. *Aid*: Let $(z_1,z_2) \in \mathcal{S} \times \mathcal{S}$ and $(x,y) \in \mathcal{S} \times \mathcal{S}$. Consider the matrix $\mathbf{P}$ with arrays

$$\mathbf{P}\left((x,y),(z_1,z_2)\right)$$

for the conditional probability of transition from $\left(X_n^{(1)}, X_n^{(2)}\right) = (x,y)$ to $\left(Z_n^{(1)}, Z_n^{(2)}\right) = (z_1,z_2)$.
Show that swapping/nonswapping preserves the distribution μ, i.e.,

$$\mu = \mu \mathbf{P}.$$

Aid: The reversibility condition

$$\mu\left((x,y)\right)\mathbf{P}\left((x,y),(z_1,z_2)\right) = \mu\left((z_1,z_2)\right)\mathbf{P}\left((z_1,z_2),(x,y)\right).$$

may turn out to be helpful. (2 p)

c) In view of the preceding we can regard $\mathbf{Z} = \{\mathbf{Z}_n = \left(Z_n^{(1)}, Z_n^{(2)}\right)\}_{n \geq 0}$ as a Markov chain in the state space $\mathcal{S} \times \mathcal{S}$ with the one-step probability transition matrix

$$\mathbf{P}^{(1)}\mathbf{P}^{(2)}\mathbf{P}.$$

Show now that

$$P\left(Z_n^{(1)} = x\right) = \pi_1(x).$$

In other words, if we record the values of the first component of $\mathbf{Z}$ we get the data that simulates π_1. (1 p)

d) When MC^3 is implemented in practice, then one usually takes $\pi_1(x) = \pi(x)$ and $\pi_2(x) \propto \pi(x)^\beta$, where $\beta = \frac{1}{1+T}$ for (the temperature) $T \geq 0$. When $T = 0$ we get the target π. Why is $\mathbf{X}^{(2)}$ a faster chain ? *Aid:* Consider what happens to the acceptance probability $\alpha_2(x,y)$ ($x =$ current state, $y =$proposed state) of $\mathbf{X}^{(2)}$ in the case where $\pi(x) > \pi(y)$, when $T = 0$ is changed to $T > 0$. (4 p)

SOLUTIONS TO THE EXAM IN SF2955 COMPUTER INTENSIVE METHODS IN MATHEMATICAL STATISTICS
WEDNESDAY 25th of may 2011 02.00 p.m. – 07.00 p.m..

Uppgift 1

The posteriori density for $\Theta_1, \Theta_2, \Theta_3$ given $\mathbf{x}^{(l)}$ for $l = 1, 2, 3, 4$, is given by Bayes' formula

$$f_{\Theta_1,\Theta_2,\Theta_3|\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)},\mathbf{x}^{(4)}}(\theta_1,\theta_2,\theta_3) = \frac{1}{c} \cdot \prod_{l=1}^{4} p_{X_1,X_2,X_3|\Theta_1,\Theta_2,\Theta_3}\left(\mathbf{x^{(l)}} \mid \theta_1,\theta_2,\theta_3\right) \cdot \prod_{i=1}^{3} \pi_{\Theta_i}(\theta_i),$$

since the samples are independent outcomes of $\mathbf{X}$ and Θ_1, Θ_2 and Θ_3 are independent random variables.
We have

$$p_{X_1,X_2,X_3|\Theta_1,\Theta_2,\Theta_3}\left(\mathbf{x^{(l)}} \mid \theta_1,\theta_2,\theta_3\right) = \prod_{i=1}^{3} p_{X_i}(x_i^{(l)} \mid \theta_i)$$

$$= \prod_{i=1}^{3} \theta_i^{x_i^{(l)}} (1-\theta_i)^{1-x_i^{(l)}}.$$

and thus

$$\prod_{l=1}^{4} p_{X_1,X_2,X_3|\Theta_1,\Theta_2,\Theta_3}\left(\mathbf{x^{(l)}} \mid \theta_1,\theta_2,\theta_3\right) = \prod_{l=1}^{4}\prod_{i=1}^{3} \theta_i^{x_i^{(l)}} (1-\theta_i)^{1-x_i^{(l)}}$$

$$= \prod_{i=1}^{3}\prod_{l=1}^{4} \theta_i^{x_i^{(l)}} (1-\theta_i)^{1-x_i^{(l)}}$$

$$= \left(\theta_1^1 \cdot (1-\theta_1)^3\right)\left(\theta_2^3 \cdot (1-\theta_1)^1\right)\left(\theta_3^3 \cdot (1-\theta_3)^1\right)$$

where we used the observed samples. The prior density is

$$\prod_{i=1}^{3} \pi_{\Theta_i}(\theta_i) = \left(\frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)}\right)^3 \cdot \prod_{i=1}^{3} \theta_i^{-1/2}(1-\theta_i)^{-1/2}.$$

Thus the denominator in Bayes' formula can be re-arranged as

$$\left(\frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)}\right)^3 \left(\theta_1^{1-1/2} \cdot (1-\theta_1)^{3-1/2}\right)\left(\theta_2^{3-1/2} \cdot (1-\theta_2)^{1-1/2}\right)\left(\theta_3^{3-1/2} \cdot (1-\theta_3)^{1-1/2}\right)$$

$$= \left(\frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)}\right)^3 \left(\theta_1^{0.5} \cdot (1-\theta_1)^{2.5}\right)\left(\theta_2^{2.5} \cdot (1-\theta_2)^{0.5}\right)\left(\theta_3^{2.5} \cdot (1-\theta_3)^{0.5}\right)$$

The normalizing constant c in the posterior density is

$$c = \int_0^1 \int_0^1 \int_0^1 \prod_{l=1}^{4} p_{X_1,X_2,X_3|\Theta_1,\Theta_2,\Theta_3}\left(\mathbf{x^{(l)}} \mid \theta_1,\theta_2,\theta_3\right) \cdot \prod_{i=1}^{3} \pi_{\Theta_i}(\theta_i) d\theta_1 d\theta_2 d\theta_3,$$

and from the above and since the expression containing the Gamma functions appears in both numerator and denominator we only need to find

$$\int_0^1 \int_0^1 \int_0^1 \left(\theta_1^{1-1/2} \cdot (1-\theta_1)^{3-1/2}\right) \left(\theta_2^{3-1/2} \cdot (1-\theta_2)^{1-1/2}\right) \left(\theta_3^{3-1/2} \cdot (1-\theta_3)^{1-1/2}\right) d\theta_1 d\theta_2 d\theta_3$$

$$= \int_0^1 \theta_1^{0.5} \cdot (1-\theta_1)^{2.5} d\theta_1 \int_0^1 \theta_2^{2.5}(1-\theta_2)^{0.5} d\theta_2 \int_0^1 \theta_3^{2.5} \cdot (1-\theta_3)^{0.5} d\theta_3.$$

This is a product of three Beta integrals and the collection of formulas in Bayesian statistics gives by the formula for the Beta integral that

$$\int_0^1 \theta_1^{0.5} \cdot (1-\theta_1)^{2.5} d\theta_1 = \frac{\Gamma(0.5)\Gamma(2.5)}{\Gamma(3)}.$$

In addition

$$\int_0^1 \theta_2^{2.5} \cdot (1-\theta_2)^{0.5} d\theta_2 = \int_0^1 \left(\theta_3^{2.5} \cdot (1-\theta_3)^{0.5}\right) d\theta_2$$

$$= \frac{\Gamma(2.5)\Gamma(0.5)}{\Gamma(3)}.$$

Thus we get

$$f_{\Theta_1,\Theta_2,\Theta_3|\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)},\mathbf{x}^{(4)}}(\theta_1,\theta_2,\theta_3)$$

$$= \begin{cases} \left(\frac{\Gamma(3)}{\Gamma(0.5)\Gamma(2.5)}\right)^3 \left(\theta_1^{0.5} \cdot (1-\theta_1)^{2.5}\right) \left(\theta_2^{2.5} \cdot (1-\theta_2)^{0.5}\right) \left(\theta_3^{2.5} \cdot (1-\theta_3)^{0.5}\right), & 0 < \theta_i < 1 \text{ for } i = 1,2,3, \\ 0 & \text{elsewhere.} \end{cases}$$

This is a product of three Beta distributions, one Beta$(1.5, 3.5)$ and two Beta$(3.5, 1.5)$s.

Uppgift 2

a)
We take $\theta = P(X > 1)$ and the plug-in estimate is $\widehat{\theta} =$ the relative frequency of samples > 1. From the data we get the numerical value $\widehat{\theta} = 6/20 = 0.3$. It holds that the corrresponding estimator $20\widehat{\theta}(X_1, X_2, \cdots, X_{20})$ has the distribution Bin$(20, \theta)$.

b)
When performing a non-parametric bootstrap we draw from the 20 observations with replacement. This is the typical situation for Binomial distribution and hence we get that $20\widehat{\theta}^*$ has the distribution Bin$(20, 0.3)$.

Uppgift 3

a) We develop the numerator in the acceptance probability

$$\alpha\left((y_1,y_2),(x_1,x_2)\right) = \min\left(1, \frac{\pi_{X_1,X_2}(y_1,y_2)\, q\left((y_1,y_2),(x_1,x_2)\right)}{\pi_{X_1,X_2}(x_1,x_2)\, q\left((x_1,x_2),(y_1,y_2)\right)}\right).$$

There the probability of proposing (x_1, x_2) from (y_1, y_2) is

$$q\left((y_1,y_2),(x_1,x_2)\right) = \begin{cases} \pi_{X_1|X_2=y_2}(x_1) & \text{if } x_2 = y_2 \\ 0 & \text{if } y_2 \neq x_2. \end{cases}$$

Then

$$\begin{aligned}
&\pi_{X_1,X_2}\left(y_1,y_2\right) q\left(\left(y_1,y_2\right),\left(x_1,x_2\right)\right)\\
&= \pi_{X_1,X_2}\left(y_1,y_2\right) \pi_{X_1|X_2=y_2}(x_1)\\
&= \pi_{X_1,X_2}\left(y_1,y_2\right) \frac{\pi_{X_1,X_2}(x_1,y_2)}{\pi_{X_2}(y_2)}\\
&= \frac{\pi_{X_1,X_2}\left(y_1,y_2\right)}{\pi_{X_2}(y_2)} \pi_{X_1,X_2}(x_1,y_2)\\
&= \pi_{X_1|X_2=y_2}\left(y_1\right) \pi_{X_1,X_2}(x_1,y_2) = \pi_{X_1|X_2=x_2}\left(y_1\right) \pi_{X_1,X_2}(x_1,x_2),
\end{aligned}$$

where we used $x_2 = y_2$. But the rightmost expression is equal to the denominator in $\alpha\left(\left(y_1,y_2\right),\left(x_1,x_2\right)\right)$.

b) Next we update X_2 with X_1 kept fixed. Thus

$$q\left(\left(x_1,x_2\right),\left(y_1,y_2\right)\right) = \begin{cases} \pi_{X_2|X_1=x_1}(y_2) & \text{if } y_1 = x_1 \\ 0 & \text{if } y_2 \neq x_2. \end{cases}$$

Uppgift 4

Let $x_1,\ldots,x_n$ be I.I.D. samples of a random variable X. The parameter of interest is $\theta = E\left[X\right]$. We take B independent bootstrap samples

$$x_1^*(b),\ldots,x_n^*(b), \quad b = 1,2,\ldots,B$$

of $x_1,\ldots,x_n$. We calculate for each bootstrap replica the arithmetic mean

$$\overline{x^*}(b) = \frac{1}{n}\sum_{i=1}^{n} x_i^*(b),$$

and then set

$$\overline{x^*} = \frac{1}{B}\sum_{b=1}^{B}\overline{x^*}(b).$$

We are interested in studying the bootstrap hypothesis and estimate the bias of $\overline{x^*}$ by

$$\text{BIAS} = \overline{x^*} - \overline{x},$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

a) By rules of expectation

$$E_{\widehat{F}_n}\left[\text{BIAS}\right] = E_{\widehat{F}_n}\left[\overline{x^*} - \overline{x}\right] = E_{\widehat{F}_n}\left[\overline{x^*}\right] - \overline{x},$$

since $\overline{x}$ is a constant, when $x_1,\ldots,x_n$ are given. It holds by rules of expectation that

$$E_{\widehat{F}_n}\left[\overline{x^*}\right] = \frac{1}{B}\sum_{b=1}^{B} E_{\widehat{F}_n}\left[\overline{x^*}(b)\right] = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n}\sum_{i=1}^{n} E_{\widehat{F}_n}\left[x_i^*(b)\right].$$

Then it holds, c.f. p. 86 of GE (compendium)

$$E_{\widehat{F}_n}\left[x_i^*(b)\right] = \sum_{r=1}^{n} x_r P\left[x_i^*(b) = x_r\right] = \sum_{r=1}^{n} x_r \frac{1}{n} = \overline{x}.$$

Thus

$$\frac{1}{B}\sum_{b=1}^{B}\frac{1}{n}\sum_{i=1}^{n} E_{\widehat{F}_n}\left[x_i^*(b)\right] = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n}\sum_{i=1}^{n}\overline{x} = \frac{1}{B}\sum_{b=1}^{B}\overline{x} = \overline{x}.$$

Hence

$$E_{\widehat{F}_n}\left[\text{BIAS}\right] = 0.$$

b) By the properties of variance, as $\overline{x}$ is a constant w.r.t. $\widehat{F}_n$, we get

$$V_{\widehat{F}_n}\left[\text{BIAS}\right] = V_{\widehat{F}_n}\left[\overline{x^*} - \overline{x}\right]$$

$$= V_{\widehat{F}_n}\left[\overline{x^*}\right].$$

By independence of the bootstrap samples and a property of variance

$$V_{\widehat{F}_n}\left[\overline{x^*}\right] = \frac{1}{B^2}\sum_{b=1}^{B}\frac{1}{n^2}\sum_{i=1}^{n} V_{\widehat{F}_n}\left[x_i^*(b)\right]$$

Then it holds, c.f. p. 86 of GE (compendium)

$$V_{\widehat{F}_n}\left[x_i^*(b)\right] = E_{\widehat{F}_n}\left[\left((x_i^*(b) - E_{\widehat{F}_n}\left[(x_i^*(b)\right)^2\right]\right]$$

and by a)

$$= E_{\widehat{F}_n}\left[\left((x_i^*(b) - \overline{x}\right)^2\right]$$

$$= \sum_{r=1}^{n}\left(x_r - \overline{x}\right)^2 P\left[x_i^*(b) = x_r\right]$$

$$= \sum_{r=1}^{n}\left(x_r - \overline{x}\right)^2 \frac{1}{n} = \widehat{\sigma}^2.$$

When we insert this in the pertinent expression above we get

$$\frac{1}{B^2}\sum_{b=1}^{B}\frac{1}{n^2}\sum_{i=1}^{n} V_{\widehat{F}_n}\left[x_i^*(b)\right] = \frac{1}{B^2}\sum_{b=1}^{B}\frac{1}{n^2}\sum_{i=1}^{n}\widehat{\sigma}^2$$

$$= \frac{1}{B^2}\sum_{b=1}^{B}\frac{1}{n}\widehat{\sigma}^2 = \frac{1}{B}\frac{1}{n}\widehat{\sigma}^2,$$

as was claimed.

c) The unconditional variance of BIAS is

$$V\left[\text{BIAS}\right] = V\left[E_{\widehat{F}_n}\left[\text{BIAS}\right]\right] + E\left[V_{\widehat{F}_n}\left[\text{BIAS}\right]\right]$$

$$= E\left[V_{\widehat{F}_n}(\text{BIAS}\right],$$

since by a) $E_{\widehat{F}_n}\left[\text{BIAS}\right]=0$. We have by b) that

$$E\left[V_{\widehat{F}_n}\left[\text{BIAS}\right]\right]=\frac{1}{B}\frac{1}{n}E\left[\widehat{\sigma}^2\right].$$

We have that

$$E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]=\sigma^2,$$

which uses the unbiased estimator of σ^2. Thus

$$E\left[\text{BIAS}\right]=\frac{1}{B}\frac{1}{n}E\left[\widehat{\sigma}^2\right]=\frac{1}{B}\frac{1}{n}\frac{(n-1)}{n}E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right],$$

which gives the desired result, or,

$$V\left[\text{BIAS}\right]=\frac{n-1}{n}\frac{1}{Bn}\sigma^2.$$

Uppgift 5

a) We write

$$T(F)=\int_a^b dF(x)=\int_{-\infty}^{\infty}I_{]a,b]}(x)dF(x)=\int_{-\infty}^{\infty}a(x)dF(x).$$

Then form any two arbitrary real numbers λ and μ

$$T(\lambda F+\mu G)=\int_{-\infty}^{\infty}a(x)d\left(\lambda F(x)+\mu G(x)\right)=\int_{-\infty}^{\infty}a(x)\lambda dF(x)+\int_{-\infty}^{\infty}a(x)\mu dG(x)$$

where we used the properties of integrals, and when we used the properties of integrals again we get

$$=\lambda\int_{-\infty}^{\infty}a(x)dF(x)+\mu\int_{-\infty}^{\infty}a(x)dG(x)=\lambda T(F)+\mu T(G),$$

which proves linearity.

b) By the linearity just shown

$$T\left((1-\epsilon)F+\epsilon\delta_t\right)=(1-\epsilon)T\left(F\right)+\epsilon T\left(\delta_t\right).$$

We use the given formula

$$T(\delta_t)=\int_{-\infty}^{\infty}a\left(x\right)d\delta_t(x)=a(t).$$

This gives

$$\frac{T\left((1-\epsilon)F+\epsilon\delta_t\right)-T\left(F\right)}{\epsilon}=\frac{(1-\epsilon)T\left(F\right)+\epsilon a(t)-T\left(F\right)}{\epsilon}$$

$$=\frac{-\epsilon T\left(F\right)+\epsilon a(t)}{\epsilon}=a(t)-T\left(F\right).$$

Therefore

$$L_T(t;F)=I_{]a,b]}(t)-\left(F(b)-F(a)\right).$$

b) The empirical influence function $\widehat{L}_T(t)$ is defined as $\widehat{L}_T(t) = L_T(t;\widehat{F}_n)$, and is the plug-in version of $L_T(t;F)$. Therefore

$$\widehat{L}_T(t) = L_T(t;\widehat{F}_n) = I_{]a,b]}(t) - \left(\widehat{F}_n(b) - \widehat{F}_n(a)\right),$$

where now

$$\widehat{F}_n(b) = \frac{\text{number of } x_i \leq b}{n}$$

Thus

$$\widehat{F}_n(b) - \widehat{F}_n(a) = \frac{\text{number of } x_i \text{ in }]a,b]}{n}.$$

c) In the delta method we assume that

$$T\left(\widehat{F}_n\right) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} L_T(X_i;F),$$

and get

$$V_F\left(T(\widehat{F}_n)\right) \approx V_F\left(T(F) + \frac{1}{n}\sum_{i=1}^{n} L_T(X_i;F)\right)$$
$$= \frac{1}{n}V_F\left(L_T(X;F)\right) = \frac{1}{n}E_F\left(L_T(X;F)^2\right),$$

Then we use

$$\widehat{\tau}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{L}_V(X_i)^2$$

and hope that

$$\sqrt{n}\frac{T\left(\widehat{F}_n\right) - T(F)}{\widehat{\tau}_n} \approx N(0,1).$$

It has been shown that if this assumption is actually true, then the bootstrap hypothesis is true for $T\left(\widehat{F}_n\right) - T(F)$.

Uppgift 6

a) Since

$$P^{(1)}\left((u,v),(x_1,x_2)\right) = \begin{cases} P^{(1)}(u,x_1) & \text{if } v = x_2 \\ 0 & \text{if } v \neq x_2 \end{cases}$$

we get for any pair (x_1,x_2)

$$\sum_{(u,v)} \mu\left((u,v)\right) P^{(1)}\left((u,v),(x_1,x_2)\right)$$

$$= \sum_{(u,v):v=x_2} \mu\left((u,v)\right) P^{(1)}(u,x_1)$$

$$= \sum_{u:v=x_2} \pi_1(u)\pi_2(v)P^{(1)}(u,x_1)$$

$$= \sum_u \pi_1(u)\pi_2(x_2)P^{(1)}(u, x_1)$$

$$= \pi_2(x_2)\sum_u \pi_1(u)P^{(1)}(u, x_1),$$

and since $P^{(1)}(u, x_1)$ is the conditional transition probability of the chain $\mathbf{X}^{(1)}$ and the Markov chain $\mathbf{X}^{(1)}$ by construction preserves π_1 we get

$$= \pi_2(x_2)\pi_1(x_1) = \mu((x_1, x_2)).$$

Thus we have shown that

$$\mu = \mu\mathbf{P}^{(1)}.$$

b) The conditional probability matrix for the swap has the arrays

$$\mathbf{P}((x, y), (z_1, z_2)) = \begin{cases} \alpha & (z_1, z_2) = (y, x) \\ 0 & (z_1, z_2) \neq (y, x) \text{ or } (z_1, z_2) \neq (x, y) \\ 1-\alpha & (z_1, z_2) = (x, y). \end{cases}$$

We show now that swapping/nonswapping preserves the distribution μ. We use the reversibility condition. Thus we start with the case $(z_1, z_2) = (y, x)$

$$\mu((x, y))\mathbf{P}((x, y), (z_1, z_2))$$

$$= \mu((x, y))\alpha((x, y), (y, x)) = \mu((x, y))\frac{\pi_1(y)\pi_2(x)}{\pi_1(x)\pi_2(y)}$$

assuming that $\pi_1(y)\pi_2(x) < \pi_1(x)\pi_2(y)$, and then

$$= \pi_1(x)\pi_2(y)\frac{\pi_1(y)\pi_2(x)}{\pi_1(x)\pi_2(y)}$$

$$= \pi_1(y)\pi_2(x)$$

$$= \pi_1(y)\pi_2(x)\cdot 1$$

and as $\pi_1(y)\pi_2(x) < \pi_1(x)\pi_2(y)$ we get

$$= \pi_1(y)\pi_2(x)\min\left(1, \frac{\pi_1(x)\pi_2(y)}{\pi_1(y)\pi_2(x)}\right)$$

$$= \mu((y, x))\mathbf{P}((y, x), (x, y)) = \mu((z_1, z_2))\mathbf{P}((z_1, z_2), (x, y)),$$

since $(z_1, z_2) = (y, x)$. Hence the reversibility has been established in the case under consideration. The case $\pi_1(y)\pi_2(x) > \pi_1(x)\pi_2(y)$ can be handled by starting with

$$\mu((z_1, z_2))\mathbf{P}((z_1, z_2), (x, y)).$$

In the cases where $\mathbf{P}((x, y), (z_1, z_2)) = 0$ the reversibility conditon is trivially true. Finally, when $\mathbf{P}((x, y), (z_1, z_2)) = 1-\alpha$, the proof is as above.
Hence the claimed preservation of probability

$$\mu = \mu\mathbf{P}$$

follows by reversibility.

c) In view of the preceding we can regard $\mathbf{Z} = \{\mathbf{Z}_n = \left(Z_n^{(1)}, Z_n^{(2)}\right)\}_{n\geq 0}$ as a Markov chain in the state space $\mathcal{S} \times \mathcal{S}$ with the one-step probability transition matrix

$$\mathbf{P}^{(1)}\mathbf{P}^{(2)}\mathbf{P}.$$

Then we get that

$$\mu\mathbf{P}^{(1)}\mathbf{P}^{(2)}\mathbf{P} = \mu\mathbf{P}^{(2)}\mathbf{P} = \mu\mathbf{P} = \mu$$

i.e., μ is the invariant probability of $\mathbf{Z}$. Therefore

$$P\left(Z_n^{(1)} = x\right) = \sum_y P\left(Z_n^{(1)} = x, Z_n^{(2)} = y\right) = \sum_y \mu\left((x,y)\right)$$

$$= \sum_y \pi_1\left(x\right)\pi_2\left(y\right) = \pi_1\left(x\right)\underbrace{\sum_y \pi_2\left(y\right)}_{=1} = \pi_1\left(x\right).$$

In other words, if we record the values of the first component of $\mathbf{Z}$ we get the data that simulates π_1.

d) Let $\pi_1(x) = \pi(x)$ and $\pi_2(x) \propto \pi(x)^\beta$, where $\beta = \frac{1}{1+T}$ for (the temperature) $T \geq 0$. By the Metropolis-Hastings theory the acceptance probability of the $\mathbf{X}^{(2)}$ with $T = 0$ is

$$\alpha^{(2)}\left(x,y\right) = \min\left(1, \frac{\pi\left(y\right)q^{(2)}\left(y,x\right)}{\pi\left(x\right)q^{(2)}\left(x,y\right)}\right)$$

and with $T > 0$

$$\alpha^{(2)}\left(x,y\right) = \min\left(1, \frac{\pi(y)^\beta q^{(2)}\left(y,x\right)}{\pi\left(x\right)^\beta q^{(2)}\left(x,y\right)}\right).$$

If now $\pi\left(x\right) > \pi\left(y\right)$ then

$$\frac{\pi\left(y\right)}{\pi\left(x\right)} < \left(\frac{\pi\left(y\right)}{\pi\left(x\right)}\right)^\beta.$$

To prove this we write first $l = \frac{\pi(y)}{\pi(x)}$. Next, if $T > 0$, then $0 < \beta < 1$. By assumption $0 < l < 1$, thus $\ln l < 0$. Then $0 < \beta < 1$ implies $\beta \cdot \ln l > \ln l$, and thus $l^\beta > l$, as was to be shown.

This means that the acceptance probability increases, when we tilt $\pi(x)$ to π^β and therefore $\mathbf{X}^{(2)}$ is a faster chain when $\beta = \frac{1}{1+T}$.

Comment: In reality one uses several altervative chains with $\beta = \frac{1}{1+T}$ of respectively increasing temperature, and swaps the values of two chains by picking them at random. This may be tuned so that the basic chain $\mathbf{X}^{(1)}$ swaps values with another chain at about 60-80% of time. The practical question is if one wins that much for heat compared to the computational cost of running several parallel chains.