Statistics (1): Estimation

Marco Banterle, Christian Robert and Judith Rousseau

Practicals

2014-2015 L3, MIDO, Université Paris Dauphine

Table des matières

1 Random variables, probability, expectation

Exercice 1.1. ***Reminders*** Let X be an integrable random vaiable. Are the following propositions true/false :

- $E(1/X) = 1/E(X)$?
- $E(X^2) \geq E(X)^2$?
- If X is symetric with respect to 0 then $E[X] = 0$?
- $E(XY) = E(X)E(Y)$?

Brief justification is expected.

Exercice 1.2. ***The multinomial distribution*** Consider a population divided into k categories according to the proportions $p_1, ..., p_k$, with $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1$. We draw n individuals with relacement. Denote by N_i the number of individuals belonging to category i, among the n individuals.

1. Determine the distribution of $(N_1, ..., N_k)$.
2. What is the marginal distribution of N_j for all $j = 1, ..., k$? Compute the expectation of N_j, $j \leq k$.
3. Compute
$$P[N_1 = n_1 | N_2 = n_2]$$

Exercice 1.3.

1. Let X be a random variable following the uniform distribution on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Determine the distribution of $Y = \tan X$.
2. Consider the same question for a general random variable X having a distribution absolutely continuous distribution with respect to Lebesgue measure, with density f, continuous everywhere (apart possibly at a finite number of points).
Apply this result to the case of an exponential random variable X with parameter $\theta > 0$.

Exercice 1.4. Let X be a real random variable with density

$$f_X(x) = \left(\frac{10}{x^2}\right) \mathbb{I}_{x>10} \, .$$

Compute $\mathbb{P}(X > 20)$, $F_X(x)$ and $\mathbb{E}(X)$.

Exercice 1.5. 1. Let Y be a real random variable and Z a random variable, independent of Y, such that

$$\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2} \, .$$

Show that the law of $X = ZY$ is symmetric and compute its characteristic function according to Φ_Y (characteristic function of Y). If the random variable Y admits the density f_Y on $\mathbb{R}$, what is the law of X ?

2. Let X be a random variable following the standard Laplace law :

$$f_X(x) = \frac{1}{2}\exp(-|x|)\,.$$

Show that, for every real t, we have that,

$$\Phi_X(t) = \frac{1}{1+t^2}\,.$$

2 bivariate random variables, change of variables and independence

Exercice 2.1. Let $X = (X_1, X_2) \in \mathbf{R}^2$ be a random Gaussian vector with density with respect to Lebesgue measure on $\mathbf{R}^2$:

$$f(x_1, x_2) = \frac{e^{-(x_1^2+x_2^2)/2}}{2\pi}, \quad \forall x_1, x_2 \in \mathbf{R}$$

Let $g : \mathbf{R} \setminus \{(0,0)\} \to \mathbf{R}^{+*} \times [0, 2\pi)$ inversible such that

$$g^{-1}(r,\theta) = (r\cos(\theta), r\sin(\theta)), \quad g(X_1, X_2) = (R, \Theta)$$

Determine the distribution of (R, Θ).

Exercice 2.2. Recall that for $p > 0$, we note

$$\Gamma(p) = \int_0^{+\infty} e^{-t}\, t^{p-1}\, dt$$

and denote $\Gamma(p,\theta)$ the distribution on $\mathbf{R}_+$ with density with respect to Lebesgue measure

$$t \mapsto \frac{\theta^p}{\Gamma(p)}\, e^{-\theta\, t}\, t^{p-1}\, \mathbf{1}_{\{t>0\}}. \tag{1}$$

We also denote, for $p, q > 0$, Beta(p, q) the Beta distribution on $[0,1]$ whosedensity with respect to Lebesgue measure is given by

$$t \mapsto \frac{1}{B(p,q)}\, t^{p-1}\, (1-t)^{q-1}\, \mathbf{1}_{\{t \in]0,1[\}}$$

whee $B(p,q)$ is a normalising constant.

Assume that X and Y are 2 independent random variables following $\Gamma(p,\theta)$ and $\Gamma(q,\theta)$ respectively.

1 Check that (**??**) defines a probability density function.
2 Derive the expectation of this law.
3 Show that the law of the sum $X+Y$ is the $\Gamma(p+q,\theta)$.
4 Derive the law of θX.

5 Show that the random variables $X+Y$ and $\frac{X}{Y}$ are independent and compute their laws.
6 Show that the random variables $X+Y$ and $\frac{X}{X+Y}$ are independent. Compute the law of the random variable $\frac{X}{X+Y}$.
7 Show that the distribution of X/Y is independent of θ and determine its density.

Exercice 2.3. ***Chi-square and Student*** Let $(X_n)_n$ be independent Gaussian random variables $\mathcal{N}(0,1)$. Define for all $n \in \mathbb{N}^*$

$$Y_n = \sum_{j=1}^{n} X_j^2, \quad T_n' = \frac{X_{n+1}}{\sqrt{Y_n}}, \quad T_n = \sqrt{n} T_n'.$$

1. Chi-square : Show by induction that the distribution of Y_n is absolutely continuous wrt Lebesgue measure with density f_{Y_n} :

$$\forall y \in \mathbf{R} \quad f_{Y_n}(y) = \mathbf{1}_{y>0} \frac{2^{-n/2}}{\Gamma(n/2)} y^{n/2-1} \exp\left(-\frac{y}{2}\right)$$

2. Compute the expectation and the variance of Y_n.
3. Student : Compute the density of the distribution of T_n' with respect to Lebesgue measure and show that the distribution of T_n has a density wrt Lebesgue measure given by

$$\forall y \in \mathbf{R} \quad f_{T_n}(y) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{y^2}{n}\right)^{-(n+1)/2}$$

4. When is T_n integrable ? T_n ?

Exercice 2.4.

Consider n random variables $X_1, X_2, \ldots, X_n$ iid from a distribution with cdf F. We assume that F is strictly monotone and continuous.

Give the distribution of $Z = -2\sum_{i=1}^{n} \log(F(X_i))$.

Exercice 2.5. Define

$$X_n = \begin{cases} n^{1/2} & U \leq 1/n \\ 0 & \text{otherwise} \end{cases} .$$

where $U \sim \mathcal{U}([0,1])$.

Study the convergence L^2, $\mathbb{P}$ et $\mathcal{L}$ of the sequence (X_n).

Exercice 2.6. 1. Compute the moment-generating function of the uniform law on $\{2, \ldots, 12\}$

2. Let X_1 and X_2 be independent random variables with values in $\{1,\dots,6\}$. Studying the roots of the polynomial $G_{X_1}G_{X_2}$ (where G_{X_i} is the moment-generating function of X_i), show that the law of the random variable X_1+X_2 cannot be the uniform law on $\{2,\dots,12\}$ (we should find that for every real s and every $i \in \{1,2\}$, there is a polynomial ϕ_i of odd degree with real coefficients such that for every strictly positive real s, we have $G_{X_i}(s) = s\phi_i(s)$).
3. Can we build two 'loaded dices' (two independent random variables) such that every outcome of their sum between 2 and 12 is equally likely ?

Exercice 2.7. Let $U_1,\dots,U_n$ be an n-sample of a $\mathcal{U}([a,b])$. Define $M_n = \max(U_1,\dots,U_n)$ and $N_n = \min(U_1,\dots,U_n)$.

1. Show that M_n converges in probability to b.
2. Show that N_n converges in L^2 norm toward a.
3. Show that $n(b-M_n)$ converges in distribution to a random variable that follow an exponential law with parameter $\frac{1}{b-a}$.

3 Standard distributions

Exercice 3.1. Consider a sequence X_i of iid Poisson variables with parameter $\lambda_i > 0$ denoted $\mathcal{P}(\lambda_i)$:

$$P[X_i = k] = e^{-\lambda_i}\frac{\lambda_i^k}{k!}, k \in \mathbb{N}$$

1. Show that, for all $p \in \mathbb{N}^*$
$$\sum_{i=1}^{p} X_i \sim \mathcal{P}(\sum_{i=1}^{p} \lambda_i)$$
2. For every $n \geq 1$ take the same iid sample $X_1, ..., X_n$ as above and denote N_n the number of zero realisations among those X_i's. Deduce the distribution of N_n.

Exercice 3.2. Let (X,Y) be a couple of real random variable with density

$$f_{(X,Y)}(x,y) = \frac{1}{756}(x^2+xy)\mathbb{I}_{0<x<6}\mathbb{I}_{0<y<6}\,.$$

Compute $f_X(x)$, $f_{Y|X=x}(y)$, $\mathbb{P}(X<Y)$ and $\mathbb{E}(Y|X=x)$.

Exercice 3.3. Let $(U_n)_{n\geq 0}$ be a sequence of independent and identically distributed random variables with Bernoulli law of parameter $p \in]0,1[$, for every positive integer n, we denote $Y_n = U_nU_{n+1}$ and $S_n = Y_1 + \cdots + Y_n$

1. For every integer $n \geq 0$, what is the law of Y_n ?

2. At what condition on the integers n and m such that $0 \leq n < m$ we get that the random variables Y_n and Y_m are independent ?
3. Compute $\mathbb{E}[Y_n Y_m]$, then compute $\mathbb{E}\left[\frac{S_n}{n}\right]$.
4. Show that there exists a real constant C such that $\mathbb{V}[S_n] \leq Cn$.
5. Establish that the sequence $\left(\frac{S_n}{n}\right)_{n\geq 0}$ converges in probability toward a constant (and specify this limit).

Exercice 3.4. 1. Let X and Y be two independent random variables with law $\mathcal{N}(m_1, \sigma_1^2)$ and $\mathcal{N}(m_2, \sigma_2^2)$ respectively, what is the law of $X + Y$?

2. Let $X_1, \ldots, X_n$ be iid random variables with distribution $\mathcal{N}(m, \sigma^2)$, what is the law of the random variable $\overline{X_n} = \frac{X_1 + \cdots + X_n}{n}$?
3. Show that the random variable $\frac{\sqrt{n}}{\sigma}(\overline{X_n} - m)$ follow the $\mathcal{N}(0, 1)$ law.
4. Define $\alpha \in]0, 1[$, show that there exists a unique positive real number ϕ_α such that,

$$\int_{-\phi_\alpha}^{\phi_\alpha} \frac{\exp(-\frac{x^2}{2})}{t} \sqrt{2\pi} dx = 1 - \alpha$$

5. Derive that exists an interval $I_\alpha = [m - t, m + t]$ with a real t such that $\mathbb{P}(\overline{X_n} \in I_\alpha) = 1 - \alpha$.
6. Show that for every real strictly positive number ϵ, we have that

$$\lim_{n\to+\infty} \mathbb{P}(|\overline{X_n} - m| > \epsilon) = 0\,.$$

Exercice 3.5. Consider a sequence $(X_n)_{n\geq 1}$ of real random variables. For $n \in \mathbb{N}^*$, X_n follow the exponential law with parameter n. Define

$$Y_n = \sin\left([X_n]\frac{\pi}{2}\right),$$

where $[X_n]$ is the integer part of X_n.

1. Find the distribution of the random variable Y_n, and compute $\mathbb{E}(Y_n)$.
2. Show that the sequence $(Y_n)_{n\geq 1}$ converges in distribution toward a constant random variable Y ; specify the law of Y .
3. Check the convergence in probability of the sequence $(Y_n)_{n\geq 1}$.

Exercice 3.6. 1. Derive the characteristic function of the uniform law in $[-1, 1]$.

2. For every n, define the random variable X_n with

$$\mathbb{P}(X_n = 1/2^n) = \mathbb{P}(X_n = -1/2^n) = \frac{1}{2}$$

and suppose that X_n are mutually independent. Define $S_n = \sum_{i=1}^n X_i$. Show that (S_n) converges in distribution toward a random variable S and precise its law.

Exercice 3.7. Normalisation of the asymptotic variance.
Let (X_n) be a sequence of i.i.d. random variables with law $\mathbb{P}$. Suppose that $\mathbb{E}(X_1^2) < \infty$ such that the Central Limit Theorem applies :

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

for $\sigma^2 = \mathrm{Var}\,(X_1) > 0$.

1. Making use of Slutsky's theorem, derive a sequence of random variables (a_n) function of $X_1, \ldots, X_n$ such that
$$\sqrt{n/a_n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$
2. Suppose now that σ^2 is a function of μ. Applying Slutsky's method, derive the function ϕ such that $\sqrt{n}(\phi(\overline{X}_n) - \phi(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.
3. Find (a_n) et ϕ in the particular cases $\mathbb{P} = \mathcal{B}(p)$ with $0 < p < 1/2$ and $\mathbb{P} = \mathcal{E}(\lambda)$ with $\lambda > 0$.

Exercice 3.8. ***Exponential random variables : memoryless variables***

Let T be a real random variable. Assume that T satisfies $P\,(T > 0) > 0$ together with the following condition :

$$\forall\,(s, t) \in R_+^{*2},\ P\,(T > t + s) = P(T > s)P(T > t)$$

1. Show that for all $t > 0$, $P(T > t) > 0$.
2. Consider the following application $f :]0, +\infty[\ \to \mathbb{R}, t \mapsto \ln P(T > t)$. Show that for all positive real $t > 0$
$$f(t) = tf(1)$$
3. Show that it implies that T is an exponential random variable.

Exercice 3.9. A system works using two different machines serially. The life expectancy X_1 and X_2 of the two machines follow the exponential distribution with parameters λ_1 and λ_2. The random variables X_1 and X_2 are supposed independent.

1. Show that
$$X \sim \mathcal{E}(\lambda) \text{ iff } \forall x > 0,\ \mathbb{P}(X > x) = \exp(-\lambda x)$$
2. Compute the probability that the system do not break down before time t. Infer the law of the survival time Z of the system.
3. Compute the probability that a breakdown is due to machine 1.
4. Let $I = 1$ if the breakdown is due to machine 1, $I = 0$ otherwise. Compute $\mathbb{P}(Z > t, I = \delta)$, for every $t \geq 0$ and $\delta \in \{0, 1\}$. Show that Z et I are independent.
5. Suppose that we have n identical systems that function independently and we observe their survival times $Z_1, ...Z_n$. Write the corresponding parametric model.

Exercice 3.10. Define the *generalised inverse* of F, F^-, by

$$F^-(u) = \inf\{x; F(x) \geq u\}$$

than we have that

If $U \sim \mathcal{U}(0,1)$, then the random variable $F^-(u)$ has the distribution of F.

(known as the Inverse Transform method)
Suppose than that you have available a random number generator from $U \sim \mathcal{U}(0,1)$.

1. How would you obtain samples from a random variable $X \sim \mathcal{E}(1)$ using U ?
2. How would you obtain samples from $Y \sim \mathcal{E}(\lambda)$, $\lambda > 0$ using X ?
3. How would you obtain samples from $Z \sim \Gamma(n, \lambda)$, $n \in \mathbb{N}$, $\lambda > 0$ using Y ?
4. How would you obtain samples from $W \sim \mathcal{B}eta(n_1, n_2)$, $n \in \mathbb{N}$?

Relate all of them to the first generator U.

4 Statistical Models : exponential families

Exercice 4.1. Which of the following distributions belong to the exponential family ?

a) Normal distribution with known mean μ ;

b) Binomial distribution $f(y|\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$, n known ;

c) Gamma distribution $f(y|\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-y\beta}$, shape parameter $\alpha > 0$ known ;

d) Gamma distribution $f(y|\alpha) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-y\beta}$, scale parameter $\beta > 0$ known ;

e) Exponential distribution $f(y|\lambda) = \lambda e^{-\lambda y}$;

f) Negative Binomial Distribution $f(y|\theta) = \binom{y+r-1}{r-1}\theta^r(1-\theta)^y$, r known ;

g) Uniform distribution $f(y|\theta) = \frac{1}{\theta}$, $0 < y < \theta$;

h) Pareto distribution $f(y|\theta) = \theta y^{-\theta-1}$;

i) Extreme value (Gumbel) distribution $f(y|\theta) = \frac{1}{\phi}\exp\left\{\frac{y-\theta}{\phi} - \exp\left\{\frac{y-\theta}{\phi}\right\}\right\}$, scale parameter $\phi > 0$ known.name

For those distributions which are part of the exponential family also tell if the distribution is in canonical form and what is the natural parameter.
Note : we say that a distribution belonging to the exponential family is in its canonical form if it is expressed according to its natural parameter space, i.e. if $\tau(\theta) = \theta$.

Exercice 4.2. Are the following families minimal exponential ? What is the natural parameter space in each case ?

a) Gamma distribution $f(y|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-y\beta}$;

b) Beta distribution $f(y|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}y^{\alpha-1}(1-y)^{\beta-1}$;

Exercice 4.3. Remember that in class you have seen that :

Proposition 1. *If* $\mathsf{T}(x) \in \mathbb{R}^d$ *and* $\mathsf{T}(X)$ *has density* $c(\theta)\exp\{\mathsf{T}(x)^T\theta\}$, *then, if* $l = l_1 + \ldots + l_d$ *:*

$$\mathbb{E}_\theta\left[\prod_{i=1}^d \mathsf{T}_i(X)^{l_i}\right] = -\frac{\partial^l}{\partial\theta_1^{l_1}\cdots\partial\theta_d^{l_d}}\log c(\theta)$$

Use this to deduce mean and variance of :

a) the Binomial distribution $f(y|\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$, with n known ;

b) the Gamma distribution $f(y|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-y\beta}$.

Exercice 4.4. ([optional] Lauritzen (1996)) Consider $\mathsf{X} = (x_{ij})$ and $\Sigma = (\sigma_{ij})$ symmetric positive-definite $m \times m$ matrices. The Wishart distribution, $\mathcal{W}_m(\alpha, \Sigma)$, is defined by the density

$$f(\mathsf{X}|\alpha, \Sigma) = \frac{|\mathsf{X}|^{\frac{\alpha-(m+1)}{2}} \exp\left(-\mathrm{tr}(\Sigma^{-1}\mathsf{X})/2\right)}{\Gamma_m(\alpha)|\Sigma|^{\alpha/2}}$$

where tr(A) is the trace of A and

$$\Gamma_m(\alpha) = 2^{\alpha m/2}\, \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\left(\frac{\alpha - i + 1}{2}\right)$$

a) Show that this distribution belongs to the exponential family ;

b) Give its canonical form ;

c) Derive the mean of $\mathcal{W}_m(\alpha, \Sigma)$;

d) Show that if $z_1, \ldots, z_n \sim \mathcal{N}_m(0, \Sigma)$, then

$$\sum_{i=1}^{n} z_i z_i^T \sim \mathcal{W}_m(n, \Sigma)$$

Exercice 4.5. Recall that the beta $\mathcal{B}e(\alpha, \beta)$ distribution has a density given by

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta}, \qquad 0 \leq \theta \leq 1\,.$$

1. Give the mean of the $\mathcal{B}e(\alpha, \beta)$ distribution.
2. Show that there a one-to-one correspondence between (α, β) and the triplet $(\mu, \theta_0, \theta_1)$, where $\pi(\theta \in [\theta_0, \theta_1]) = p$ and μ is the mean of the distribution.
3. What are the conditions on $(\mu, \theta_0, \theta_1)$ for (α, β) to exist ?

Exercice 4.6. Dynkin (1951)
Show that the normal distributions and distributions of the form $c\log(y)$, when $y \sim \mathcal{G}(\alpha, \beta)$, are the only ones which can belong simultaneously to an exponential family and a location family. Deduce that the normal distribution is the only distribution from an exponential family that is also spherically symmetric.

Exercice 4.7. * Morris(1982)
A *restricted natural exponential* family on $\mathbb{R}$ is defined by

$$P_\theta(x \in A) = \int_A \exp\{\theta x - \psi(\theta)\}\, dF(x), \qquad \theta \in \Theta. \tag{2}$$

a. Show that, if $0 \in \Theta$, F is necessarily a cumulative distribution function. Otherwise, show that the transformation of F into

$$dF_0(x) = \exp\{\theta_0 x - \psi(\theta)\}\, dF(x),$$

for an arbitrary $\theta_0 \in \Theta$ and the replacement of θ by $\theta - \theta_0$, provides this case.

b. Show that, in this restricted sense, $\mathcal{B}e(m\mu, m(1-\mu))$ and the lognormal distribution $\mathcal{L}og\mathcal{N}(\alpha, \sigma^2)$ do not belong to an exponential family.
c. If $\mu = \psi'(\theta)$ is the mean of the distribution (**??**), the *variance function* of the distribution is defined by $V(\mu) = \psi''(\theta) = \mathbb{V}\mathrm{ar}_\theta(x)$. Show that V is indeed a function of μ and, moreover, that if the variation space of μ, Ω, is known, the couple (V, Ω) completely characterizes the family (**??**) by

$$\psi\left(\int_{\mu_0}^{\mu} \frac{dm}{V(m)}\right) = \int_{\mu_0}^{\mu} \frac{m\, dm}{V(m)}.$$

(Notice that $\theta = \int_{\mu_0}^{\mu} dm/V(m)$.) Show that $V(\mu) = \mu^2$ defines *two* families, depending on whether $\Omega = \mathbb{R}^-$ or $\Omega = \mathbb{R}^+$.
d. Show that $V(\mu) = \mu(1-\mu)/(m+1)$ corresponds simultaneously to the binomial distribution $\mathcal{B}(m, \mu)$ and to $\mathcal{B}e(m\mu, m(1-\mu))$. Deduce that the characterization by V is only valid for natural exponential families.
e. Show that exponential families with *quadratic variance functions*, i.e.,

$$V(\mu) = v_0 + v_1\mu + v_2\mu^2, \tag{3}$$

include the following distributions : normal, $\mathcal{N}(\mu, \sigma^2)$, Poisson, $\mathcal{P}(\mu)$, gamma, $\mathcal{G}(r, \mu/r)$, binomial, $\mathcal{B}(m, m\mu)$, and negative binomial, $\mathcal{N}eg(r, p)$, defined in terms of the number of successes before the rth failure, with $\mu = rp/(1-p)$.
f. Show that the normal distribution (respectively, the Poisson distribution) is the unique natural exponential distribtion with a constant (reap., of degree one) variance function.
g. Assume $v_2 \neq 0$ in (**??**) and define $d = v_1^2 - 4v_0v_2$, discriminant of (**??**), $a = 1$ if $d = 0$ and $a = \sqrt{dv_2}$ otherwise. Show that $x^* = aV'(x)$ is a linear transformation of x with the variance function

$$V^*(\mu^*) = s + v_2(\mu^*)^2, \tag{4}$$

where $\mu^* = aV'(\mu)$ and $s = -\mathrm{sign}(dv_2)$. Show that it is sufficient to consider V^* to characterize natural exponential families with a quadratic variance function, in the sense that other families are obtained by inverting the linear transform.
h. Show that (**??**) corresponds to six possible cases depending on the sign of v_2 and the value of s $(-1, 0, 1)$. Eliminate the two impossible cases and identify the families given in e, above. Show that the remaining case is $v_2 > 0$, $s = 1$. For $v_2 = 1$, this case corresponds to the distribution of $x = \log\{y/(1-y)\}/\pi$, where

$$y \sim \mathcal{B}e\left(\frac{1}{2} + \frac{\theta}{\pi}, \frac{1}{2} - \frac{\theta}{\pi}\right), \qquad |\theta| < \frac{\pi}{2},$$

and

$$f(x|\theta) = \frac{\exp[\theta x + \log(\cos(\theta))]}{2\cosh(\pi x/2)}. \tag{5}$$

(The reflection formula $B(0.5+t, 0.5-t) = \pi/\cos(\pi t)$ can be of use.) The distributions spanned by the linear transformations of (**??**) are called $\mathrm{GHS}(r, \lambda)$ (meaning *generalized hyperbolic secant*, with $\lambda = \tan(\theta)$, $r = 1/v_2$, and $\mu = r\lambda$. Show that the density of $\mathrm{GHS}(r, \lambda)$ can be written

$$f_{r,\lambda}(x) = \left(1 + \lambda^2\right)^{-r/2} \exp\{x \arctan(\lambda)\} f_{r,0}(x)$$

(do not try to derive an explicit expression for $f_{r,0}$).

5 Glivenko-Cantelli, empirical cumulative distribution function and Bootstrap

Exercice 5.1. Let $X_1, \ldots, X_n \sim F$ and let $\hat{F}_n(x)$ be the empirical distribution function. For a fixed x, find the limiting distribution of $\hat{F}_n(x)$. (hint : central limit theorem)

Exercice 5.2. Let $x_1, \cdots, x_n$ be n independent realisations from the same distribution P on $\mathbf{R}$. Let $A_1, \cdots, A_k$ form a partition of $\mathbf{R}$, i.e. $A_j \cup A_i = \emptyset$ if $i \neq j$ and $\cup_{i=1}^k A_i = \mathbf{R}$ and assume that each A_j is an interval.

1 Consider the empirical distribution defined as

$$P_n = \frac{1}{n}\sum_{i=1}^n \delta_{(x_i)}, \quad \text{i.e.} \quad P_n(A) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i \in A}$$

Show that for all $j \leq k$, almost surely with respect to P,

$$\lim_n P_n(A_j) = P(A_j)$$

2 Show that

$$\sqrt{n}(P_n(A_j) - P(A_j)) \Rightarrow \mathcal{N}(0, v_j), \quad \forall j \leq k$$

and determine v_j.

3 Determine the limiting distribution of the vector :

$$\sqrt{n}(P_n(A_1) - P(A_1), \cdots, P_n(A_k) - P(A_k))$$

Is it absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^k$?

Exercice 5.3. We want to estimate the cumulative distribution function F of the survival time T of a mouse who had been injected a given dose of pathogen. For obvious reasons, it is important to kill as little mice as possible during the experiment, while retaining a good precision of our estimate. What would be a *good* sample size n so that

$$\mathbb{P}[\sup_t |F_n(t) - F(t)| > 0.05] < 0.001?$$

The probability $\mathbb{P}$ is computed under the distribution of a n sample of independent and identically distributed mice whose survival time has cumulative distribution function F.

Exercice 5.4. Consider an iid sample Y_1, Y_2, Y_3 of size $n = 3$. For simplicity in thinking about the problem, suppose that Y_i are continuous so that all three are distinct with probability 1.

a) Consider drawing samples of size $n = 3$ with replacement from the set $\{Y_{(1)}, Y_{(2)}, Y_{(3)}\}$, where $Y_{(1)} < Y_{(2)} < Y_{(3)}$. Write down the $3^3 = 27$ equally likely resamples samples.

b) How many distinct resamples are obtained ? Find the general number for a sample size of n.

c) Using the samples obtained in a), devise the bootstrap distribution of the sample median.

d) From c), write down the bootstrap expectation and variance of the sample median.

Exercice 5.5. Let $X_1, \ldots, X_n$ be distinct observations (no ties).
Let $X_1^*, \ldots, X_n^*$ denote a bootstrap sample and let $\bar{X}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$.
Find : $\mathbb{E}(\bar{X}_n^*|X_1, \ldots, X_n)$, $\mathbb{V}\mathrm{ar}(\bar{\mathrm{X}}_\mathrm{n}^*|\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{n})$, $\mathbb{E}(\bar{X}_n^*)$ and $\mathbb{V}\mathrm{ar}(\bar{\mathrm{X}}_\mathrm{n}^*)$

Exercice 5.6. Suppose a *nonparametric* bootstrap sample of size n, drawn uniformly with replacement from $x_1, \ldots, x_n$ contains j_1 copies of x_1, j_2 copies of x_2 and so on till j_n copies of x_n, with $\sum_{i=1}^{n} j_n = n$.
Show that the probability of of obtaining this sample is

$$\binom{n}{j_1 j_2 \cdots j_n} \prod_{i=1}^{n} n^{-j_i} \qquad \text{where} \qquad \binom{n}{j_1 j_2 \cdots j_n} = \frac{n!}{j_1! j_2! \cdots j_n!}$$

What is the probability of obtaining that sample with the *parametric* bootstrap ?

Exercice 5.7. * Let $T_n = \bar{X}_n^2$, $\mu = \mathbb{E}(X_1)$, $\alpha_k = \int |x-\mu|^k \mathrm{dF(x)}$ and $\hat{\alpha}_k = \sum_{i=1}^{n} |X_i - \bar{X}_n|^k$.
Show that

$$\mathbb{V}\mathrm{ar}_{boot} = \frac{4\bar{X}_n^2 \hat{\alpha}_2}{n} + \frac{4\bar{X}_n \hat{\alpha}_3}{n^2} + \frac{\hat{\alpha}_4}{n^3}$$

6 Monte Carlo

Exercice 6.1. Let $\mathbf{X} = (X_1, X_2)$ be uniformly distributed on the square $[-1,1] \times [-1,1]$. Let $Y = f(\mathbf{X})$ where

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } x_1^2 + x_2^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Imagine using a Monte Carlo sample of size n = 1000 from $\mathbf{X}$

a) Detail how to produce an estimate of $\mu = \mathbb{E}(Y)$;

b) Translate your estimator in an estimate for π ;

c) Give an expression for the variance of $\hat{\pi}$;

Now imagine to have a sample from $\tilde{\mathbf{X}} = (\tilde{X}_1, \tilde{X}_2)$, uniformly distributed on the unit square $[0,1] \times [0,1]$. Can you derive a similar estimator for π ? Would this be an improvement with respect to the previous one ? Why ?

Exercice 6.2. Let $\mu_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ (the empirical mean) and $S_n = \sum_{i=1}^{n}(x_i - \mu_n)^2$ (n times the empirical variance).

There is a way to obtain good numerical stability in updating these quantities as new data points are available : Starting with $\mu_1 = x_1$ and $S_1 = 0$, make the updates as :

$$\begin{aligned}\delta_i &= x_i - \mu_{i-1}\\ \mu_i &= \mu_{i-1} + \delta_i/i\\ S_i &= S_{i-1} + \frac{i-1}{i}\delta_i^2\end{aligned}$$

for $i = 2, \ldots, n$.

Prove that this methods actually yields the expected μ_n and S_n.

Exercice 6.3. Suppose you have a sample of size $n = 1000$ from a distribution F. How would you compute an approximate value of $F(x)$ for $x = 2$? What is the variance of your estimate ? Now suppose F is a standard normal random variable. How many simulated random samples are needed to obtain three digits of accuracy for your estimator ?

Exercice 6.4. Take a Cauchy $\mathcal{C}(0, 1)$ random variable X with density

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

Show that the $\mathcal{C}(0, 1)$ has no mean.

What is the consequence on $\bar{X}_n = \sum_{i=1}^{n} X_i$, the Monte Carlo estimator for its mean ?

Exercice 6.5. We want to evaluate the integral

$$I = \int_{2}^{\infty} \frac{1}{\pi(1+x^2)}\mathrm{dx}$$

a) Analytically calculate the value of I.

b) Devise a Monte Carlo Estimator $\hat{I}_n$ for I, based on a sample of n observations from a distribution F of your choice.

* Can we do better ? Try to devise a second estimator based on a transformation of I.

Exercice 6.6. Let $N \sim \mathcal{N}(\theta, 1)$ ($\theta \in \mathbb{R}$), $X = \exp(N)$ and suppose to have an n-sample $(X_1, \ldots, X_n)$ from X.

a) Show that the density of X is

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}x}\exp\left\{-\frac{1}{2}(\log(x) - \theta)^2\right\}\mathbb{I}_{x>0};$$

b) Compute $\mathbb{E}_\theta(X)$ (hint : everything is easier using the change of variable $y = \log(x)$) ;

c) Let $T_n = \log\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \frac{1}{2}$. Show that T_n converges in probability towards θ.

7 Likelihood and Maximum Likelihood Estimation

Exercice 7.1. Let X be a random variable admitting the following probability density :

$$f_X(x;\theta) = \frac{\theta}{x^{\theta+1}} \mathbb{I}_{x \geq 1}$$

where $\theta > 0$. It is in fact a particular Pareto law. Consider a sample of size n : $X_1, \ldots, X_n$ iid from X.

1 Show that the model belongs to the exponential family and deduce that is *regular*.

2 Compute the Fisher information contained in X for the parameter θ. Deduce the information contained in the n-sample.

3 Actually we do not observe X, but a random variable Y defined by :

$$Y = \begin{cases} 1 & \text{if} \quad X \geq \exp(1) \\ 0 & \text{if} \quad X < \exp(1) \end{cases} .$$

Compute the Fisher information brought by Y for the parameter θ.

4 Show that $I_X(\theta) > I_Y(\theta)$ (we can use the fact that $\exp(x) - 2x > 0$ for every $x \geq 0$).

Exercice 7.2. We consider a random variable X following an exponential law with parameter $\theta > 0$:

$$f_X(x;\theta) = \theta \exp(-\theta x) \mathbb{I}_{x>0} .$$

Let $X_1, \ldots, X_n$ be an n-sample from X.

1 Show that $\sum_{i=1}^n X_i$ is a sufficient and minimal statistics for θ.

2 Admit that $g(\underline{X}) = \frac{X_n}{\sum_{i=1}^n X_i}$ is an ancillary statistics for θ.
Compute $\mathbb{E}(g(\underline{X}))$.

Exercice 7.3. Consider a random variable X with density

$$f_X(x;\theta) = kx^\theta \mathbb{I}_{]0,1]}(x) .$$

Let $X_1, \ldots, X_n$ be an n-sample from X.

1 Determine k as a function of θ, specifying the conditions on θ. Compute $\mathbb{E}_\theta(X)$ and $\mathbb{V}_\theta(X)$.

2 Show that $-\log(X)$ follows a known law and specify its parameter. Deduce that $\tilde{\theta}_n = -1-(n-1)\left(\sum_{i=1}^{n}\log(X_i)\right)^{-1}$ is an unbiased estimator for θ.

Exercice 7.4. Let X be a random variable admitting the following probability density :

$$f_X(x;\theta) = \frac{\theta\exp(\theta x)}{\exp(\theta^2)-1}\mathbb{I}_{[0,\theta]}(x)$$

where $\theta > 0$. Consider an n-sample $X_1,\ldots,X_n$ from X. Find, if it exists, a sufficient statistics for θ.

Exercice 7.5. Let X be a random variable with geometric law with parameter $p \in]0,1[$ (number of trials before the first success, p success probability) and $X_1,\ldots,X_n$ be an n-sample from X. Assume $p = 1-q$.

1 Compute the Fisher information brought by X on p and the one contained in the n-sample.

2 Show that $\overline{X_n}$ is sufficient and that $\overline{X_n}$ is an effective estimator for the parameter q/p.

Exercice 7.6. Let X be a Poisson random variable with parameter $\theta > 0$ and $X_1,\ldots,X_n$ an n-sample from X.

1 Show that $\overline{X_n}$ and S^2_{n-1} are unbiased estimators of θ.

2 Show that $\overline{X_n}$ is the uniformly minimum-variance unbiased estimator of θ. Deduce that $\mathbb{V}_\theta(\overline{X_n}) \leq \mathbb{V}_\theta(S^2_{n-1})$.

Exercice 7.7. Let X be a random variable in the set of the reals, admitting the density

$$f_X(x;\theta) = k\exp(-\theta|x|)$$

where $\theta > 0$ and $X_1,\ldots,X_n$ an n-sample from X.

1 Determine the constant k.

2 Compute the integrals $\int_{-\infty}^{0} t\exp(\theta t)dt$ and $\int_{0}^{+\infty} t\exp(-\theta t)dt$ then deduce the expressions of $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$.

3 Compute the estimator W_n of θ via the maximum likelihood method.

4 Show that W_n is convergent estimator for θ in quadratic mean.

Exercice 7.8. Consider an n-sample $(X_1, \ldots, X_n)$ iid from X, with density :

$$f_X(x;\theta) = \frac{3}{(x-\theta)^4}\mathbb{I}_{[\theta+1,+\infty[}\,.$$

where $\theta > 0$ is an unknown parameter.

1 Compute $\mathbb{E}_\theta(X)$ and $\mathbb{V}_\theta(X)$ (we can compute $\mathbb{E}_\theta((X-\theta))$ and $\mathbb{V}_\theta((X-\theta))$).

2 Give the maximum likelihood estimator $\hat{\theta}_n$ for θ.

3 Compute $\mathbb{E}_\theta\left(\hat{\theta}_n\right)$. Deduce an unbiased estimator $\theta_n^{\#}$ for θ as a function of $\hat{\theta}_n$.

Exercice 7.9. Let X be a random variable, distributed according to a Pareto law with parameters $\alpha > 0$ and $\beta > 0$:

$$f_X(x;\alpha,\beta) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}\mathbb{I}_{x\geq\beta}\,.$$

Let $X_1, \ldots, X_n$ be an n-sample from X. In the following $\theta = (\alpha, \beta)$.

1 Give the density of X, then compute $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$ giving the conditions on the existence on those moments.

2 Suppose β known.

a) Write the likelihood on the sample and give a sufficient statistics for α.

b) Find an estimator T_n of α with the maximum likelihood method.

c) Find the law of the random variable $Y = \log(X/\beta)$.

d) Show that T_n is a strongly consistent estimator of α.

e) Find the law of $Z_n = \sum_{i=1}^n \log(X_i/\beta)$. Deduce the expression of $\mathbb{E}(T_n)$ and $\mathbb{V}(T_n)$, then show that T_n is a quadratic mean convergent estimator for α.

f) Deduce from T_n an unbiased estimator T_n^* for α. Show that T_n^* is asymptotically efficient.

3 Suppose α known.

a) Find an estimator W_n of β via maximum likelihood.

b) Find the law of W_n and deduce that W_n is a quadratic mean convergent estimator of β.

Exercice 7.10. Consider a random variable X with density

$$f_X(x;\theta) = k|x| \exp\left(-\frac{x^2}{2\theta}\right)$$

where $\theta > 0$. Let $X_1, \ldots, X_n$ be as usual an n-sample from X.

1 Compute the normalising constant k.

2 Compute $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$.

3 Give the maximum likelihood estimator $\hat{\theta}_n$ of θ. Is it unbiased?
Is it strongly consistent?

5 Explain why the model for X is regular.

6 Compute the Fisher information given by X on θ, then the one given by the whole sample.

7 Compute $\mathbb{E}_\theta(X^4)$. $\hat{\theta}_n$ is efficient? (Cramer-Rao bound)

8 Is $\hat{\theta}_n$ the unique unbiased estimator with uniformly minimum variance?

Exercice 7.11. Let X be a real random variable with density
$f_X(x;\theta) = \frac{2\sqrt{\theta}}{\sqrt{\pi}} \exp(-\theta x^2) \mathbb{I}_{]0,+\infty[}(x)$ where $\theta > 0$.
Let $X_1, \ldots, X_n$ then be an n-sample of X.

1 Compute $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$.

2 Find W_n, the maximum likelihood estimator for θ.

Exercice 7.12. Let X be a discrete random variable with values in $\{-1, 0, 1\}$ such that $\mathbb{P}(X = 0) = 1 - 2\theta$ et $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \theta$. Suppose $\theta \in [0, 1/2]$. Consider $X_1, \ldots, X_n$ an n-sample from X.

Name R the random variable euql to the number of X_i with a non-null value. Find the maximum likelihood estimator W_n of θ. Give the law of R and deduce $\mathbb{E}(W_n)$ and $\mathbb{V}(W_n)$.

Exercice 7.13. Let X be a random variable with values in $[-1, 1]$ with density $f_X(x; a, b) = a\mathbb{I}_{[-1,0]}(x) + b\mathbb{I}_{]0,1]}(x)$ where $a \leq 0$ and $b \leq 0$ and $X_1, \ldots, X_n$ is an n-sample from X.

1 Point out the relation between a and b.
in the following, we will write b as a function of a.

2 Compute $\mathbb{E}_a(X)$ and $\mathbb{V}_a(X)$.

3 Find the maximum likelihood estimator W_n of θ.

Exercice 7.14. Infection Markers

N infectious agents aggress simultaneously an organism that has Q defending agents . The Immune response is modeliez in the following way : every defending agent choose randomly an infectious agent (only one) in the N aggressors, independently from the other defendants. With probability $\vartheta \in (0,1)$ the infectious agent is nullified.

Only one surviving infectious agent is required for the organism to be infected.

1. Show that the probability that a given aggressive agent infect the the organism is

$$p_{Q,N}(\vartheta) = \Big(1 - \frac{\vartheta}{N}\Big)^Q.$$

In the lab, we repeat n independent scenarios of aggression. In every experiment, an infectious agent is **marked**. For experiment i, we note $X_i = 1$ if the given agent did infect the organism, 0 otherwise.

2. Consider having a sample $(X_1, \dots, X_n)$, where ϑ is the unknown parameter and Q et N are known. Show that the likelihood can be written as

$$\vartheta \rightsquigarrow p_{Q,N}(\vartheta)^{\sum_{i=1}^n X_i}\big(1 - p_{Q,N}(\vartheta)\big)^{n - \sum_{i=1}^n X_i}.$$

3. Show that the model is regular and that its Fisher information is given by

$$\mathbb{I}(\vartheta) = \frac{\big(\partial_\vartheta p_{Q,N}(\vartheta)\big)^2}{p_{Q,N}(\vartheta)\big(1 - p_{Q,N}(\vartheta)\big)}.$$

4. Show that the maximum likelihood estimator for ϑ is well defined, asymptotically normal and compute its limiting variance.

Suppose now that N et Q are unknown parameters of interest, and we take the limit $N \approx +\infty$ supposing that $Q = Q_N \sim \kappa N$ for a $\kappa > 0$ (unknown).

6. Going at the limit for Nin the previous model, show that the observation of $(X_1, \dots, X_n)$ allow (identifiability) the estimate of $\widetilde{\vartheta} = \kappa\vartheta$ and hence compute its maximum likelihood estimator.

8 Bayesian Estimation

Exercice 8.1. Assume that we have perform n *independent* experiments where each experiment has probability π for success. Let $x \in \{0, 1, ..., n\}$ denote the random number of successes. The number of successes follows a Binomial distribution :

$$x \sim \mathcal{B}in(n, p)$$

that is

$$f(x|n, \pi) = \binom{n}{x} p^x (1-p)^{n-x}.$$

For a Bayesian analysis we now need to specify our prior distribution. It turns out to be convenient to specify the prior distribution for p by a beta distribution with parameter α and β, i.e.

$$p \sim \mathcal{B}e(\alpha, \beta) \text{ such that } \pi? = B(\alpha, \beta) p^{\alpha-1}(1-p)^{\beta-1} \text{ for } p \in (0, 1)$$

- Compute the posterior distribution of p
- Compute $\mathbb{E}^\pi(p)$ with respect to the prior and $\mathbb{E}^\pi(p|x)$ the posterior mean.
- Similarly, how does the posterior variance change with respect to the prior variance ?
- Remember that if $\alpha = \beta = 1$ $p(\pi|\alpha, \beta)$ is flat. Is this a non-informative prior ? If yes in which sense ?

Exercice 8.2. Consider an observation from a Normal distribution $x \sim \mathcal{N}(\mu, \sigma^2)$ where the precision $\tau = 1/\sigma^2$ is known and the mean μ is unknown. We assume a priori for μ a Normal distribution with mean μ_o and precision τ_o.

- Give the posterior distribution for μ ; what is your interpretation of the parameter of the posterior ?
- Now, the data set consists of n iid observation $\mathbf{x}^n$ from a $\mathcal{N}(\mu, \sigma^2)$ (remember that using the precision τ is probably convenient). What are the posterior mean and variance ?
- Relate the above to the maximum likelihood estimator for μ.
- What happens to the posterior distribution for $\mu|\mathbf{x}^n$ when $n \to \infty$

Give the expression of the Jeffreys Prior for this model. What is the associated posterior distribution ? What is its interpretation in terms of the conjugate prior we just described ?

Exercice 8.3. Let again $\mathbf{x}^n$ be a sample $\mathbf{x}^n = (x_1, \ldots, x_n) \sim \mathcal{N}(\mu, \sigma^2 = \frac{1}{\tau})$ but now consider that μ is known and the parameter of interest is τ. Assume a $\Gamma(\alpha, \beta)$ as a prior distribution for τ.

1

- Derive the posterior distribution $p(\tau|\mathbf{x}^n)$;
- Compute posterior mean and posterior variance for τ.

2 If both mean and variance are unknown we ideally want a joint conjugate prior distribution. Consider that we define $\pi(\mu, \tau) = \pi_1(\mu) \times \pi_2(\tau)$ where both the priors are defined as before as $\mu \sim \mathcal{N}(\mu_0, 1/\tau_0)$ and $\tau \sim \Gamma(\alpha, \beta)$. What is the marginal posterior distribution of μ ?

Exercice 8.4. Samples are taken from twenty wagonloads of an industrial mineral and analysed. The amounts in ppm (parts per million) of an impurity are found to be as follows.

```
44.3 50.2 51.7 49.4 50.6 55.0 53.5 48.6 48.8 53.3
59.4 51.4 52.0 51.9 51.6 48.3 49.3 54.1 52.4 53.1
```

We regard the observations to be drawn from a normal distribution with a known precision $\tau = 1/\sigma^2 = 0.1$ and unknown mean μ.
Compute the posterior mean, variance and 95% credibility interval for μ.

Exercice 8.5. Consider a $x \sim \mathcal{P}(\lambda)$.
- Give a conjuguate family of prior distributions for λ .
- Derive the associated posterior distributions

From the book : *The Bayesian Choice*

Exercice 8.6. If $\psi(\theta|x)$ is a posterior distribution associated with $f(x|\theta)$ and a (possibly improper) prior distribution π, show that

$$\frac{\psi(\theta|x)}{f(x|\theta)} = k(x)\pi(\theta).$$

a. Deduce that, if f belongs to an exponential family, the posterior distribution also belongs to an exponential family, whatever π is.
b. Show that if ψ belongs to an exponential family, the same holds for f.

Exercice 8.7. A *contingency table* is a $k \times \ell$ matrix such that the (i, j)-th element is n_{ij}, the number of simultaneous occurrences of the ith modality of a first characteristic, and of the jth modality of a second characteristic in a population of n individuals $(1 \le i \le k, 1 \le j \le \ell)$. The probability of this occurrence is denoted by p_{ij}.
a. Show that these distributions belong to an exponential family.
b. Determine the distributions of the margins of the table, i.e., of $n_{i\cdot} = n_{i1} + \ldots + n_{i\ell}$ and $n_{\cdot j} = n_{1j} + \ldots + n_{kj}$. Deduce the distributions of $(n_{1\cdot}, \ldots, n_{k\cdot})$ and of $(n_{\cdot 1}, \ldots, n_{\cdot \ell})$.
c. Derive conjugate priors on $p = (p_{ij})$ and the Jeffreys prior.
Recall the Jeffreys prior is defined as

$$p(\theta) \propto \sqrt{|\mathcal{I}(\theta)|}$$

where $\mathcal{I}(\theta)$ is the Fisher Information.
d. In the particular case of *independence* between the two variables, the parameters are supposed to satisfy the relations $p_{ij} = p_{i\cdot}p_{\cdot j}$ where $(p_{1\cdot}, \ldots, p_{k\cdot})$ $(p_{\cdot 1}, \ldots, p_{\cdot \ell})$ are two vectors of probabilities. Relate these vectors to the distributions derived in b. and construct the corresponding conjugate priors.