Computational Statistics

Homework 1

November 4, 2016

1.4 Integral calculation

Assume $\alpha > 0$ and $\beta > 0$.
Using Integration by substitution with $y = x^\alpha$ (this function is C^1), we get $dy = \alpha x^{\alpha-1}dx$ and:
$\int_w^{+\infty} \alpha\beta x^{\alpha-1} e^{-\beta x^\alpha} dx = \int_{w^\alpha}^{+\infty} \beta e^{-\beta y} dy$
$\int_w^{+\infty} \alpha\beta x^{\alpha-1} e^{-\beta x^\alpha} dx = [-e^{-\beta y}]_{w^\alpha}^{+\infty}$
$\boxed{\int_w^{+\infty} \alpha\beta x^{\alpha-1} e^{-\beta x^\alpha} dx = e^{-\beta w^\alpha}}$

1.7 Normal MLE

Assume $Y_1, ..., Y_n \sim N(\mu, \sigma^2) iid$. The likelihood of the model is :
$L(y_1, ..., y_n|\mu, \sigma) = \prod_{i=1}^n f(y_i|\mu, \sigma)$ where $f(y|\mu, \sigma)$ can be written as $\frac{1}{\sqrt{2\pi}} exp(-log(\sigma) + (\mu/\sigma^2)y - (1/2\sigma^2)y^2 - \mu^2/2\sigma^2)$
So by replacing $\theta_1 = \mu/\sigma^2$ and $\theta_2 = -\frac{1}{2\sigma^2}$:
$L(y_1, ..., y_n|\mu, \sigma) = (\frac{1}{\sqrt{2\pi}})^n \prod_{i=1}^n exp(\frac{1}{2} log(-2\theta_2) + \theta_1 y_i + \theta_2 y_i^2 + \frac{\theta_1^2}{4\theta_2})$
The log-likelihood is thus :
$l(y_1, ..., y_n|\mu, \sigma) = -\frac{n}{2} log(2\pi) + \frac{n}{2} log(-2\theta_2) + \sum_{i=1}^n (\theta_1 y_i + \theta_2 y_i^2 + \frac{\theta_1^2}{4\theta_2})$

The likelihood equations using the natural parameters θ_1 and θ_2 are given by :
$\begin{cases} \frac{\partial l}{\partial \theta_1}(y|\theta) = 0 \\ \frac{\partial l}{\partial \theta_2}(y|\theta) = 0 \end{cases}$
$\begin{cases} \sum_{i=1}^n y_i + \frac{2n\theta_1}{4\theta_2} = 0 \\ \frac{n}{2} \frac{-2}{-2\theta_2} + \sum_{i=1}^n y_i^2 - \frac{n\theta_1^2}{4\theta_2^2} = 0 \end{cases}$

Which gives exactly the (1.11) equations: $\begin{cases} \sum_{i=1}^n y_i = -n\frac{\theta_1}{2\theta_2} \\ \sum_{i=1}^n y_i^2 = -\frac{n}{2\theta_2} + \frac{n\theta_1^2}{4\theta_2^2} \end{cases}$

Now we can replace θ_1 and θ_2 by μ/σ^2 and $-\frac{1}{2\sigma^2}$: $\begin{cases} n\overline{y} = n\mu \\ n(s^2 + \overline{y}^2) = n(\mu^2 + \sigma^2) \end{cases}$

We therefore get the standard statistics : $\begin{cases} \hat{\mu} = \overline{y} \\ \hat{\sigma^2} = s^2 \end{cases}$

1.8 (Titterington et al. 1985) Mixture of two exponential distributions

Let $X \sim \pi Exp(1) + (1-\pi) Exp(2)$.

Then, the overall density of X can be written as :

$$f(x) = \pi e^{-x} + (1-\pi) 2e^{-2x}$$

From this, we deduce that :

$$E(X^s) = \int_0^{+\infty} x^s(\pi e^{-x} + (1-\pi)2e^{-2x})dx$$

Resulting in :

$$E(X^s) = \pi E(Y_1^s) + (1-\pi)E(Y_2^s)$$

Where $Y_1 \sim Exp(1)$ and $Y_2 \sim Exp(2)$

Let $Y \sim Exp(\lambda)$. Then, we have :

$$E(Y^s) = \int_0^{+\infty} y^s \lambda e^{-\lambda y} dy$$

Using Integration by Substitution with $u = \lambda y$:

$$E(Y^s) = \lambda^{-s} \int_0^{+\infty} u^s e^{-u} du$$

We recognize the Gamma function :

$$E(Y^s) = \lambda^{-s}\Gamma(s+1)$$

In particular :

$$E(Y_1^s) = \Gamma(s+1)$$
$$E(Y_2^s) = 2^{-s}\Gamma(s+1)$$

So :

$$\boxed{E(X^s) = (\pi + (1-\pi)2^{-s})\Gamma(s+1)}$$

Now we try to estimate π using the s moment estimator.
Let $m_s = E(X^s)$. We get :

$$m_s = (2^{-s} + \pi(1-2^{-s}))\Gamma(s+1)$$
$$(\frac{m_s}{\Gamma(s+1)} - 2^{-s})\frac{1}{1-2^{-s}} = \pi$$

As we can estimate m_2 with $\hat{m}_2 = \frac{1}{n}\sum_{i=1}^n X_i^s$ where $X_1, ..., X_n$ iid following the same distribution as X, we obtain moment estimators of π for each considered s :

$$\hat{\pi}_s = (\frac{\hat{m}_s}{\Gamma(s+1)} - 2^{-s})\frac{1}{1-2^{-s}}$$

Let $t_s(x) = \frac{x^s}{\Gamma(s+1)}$. We get : $\boxed{\hat{\pi}_s = (\frac{1}{n}\sum_{i=1}^n t_s(X_i) - 2^{-s})\frac{1}{1-2^{-s}}}$

Now we need to choose the best one (in s).
Following the law of large numbers, they are all consistent estimators of π. However, they do not have the same asymptotic variance. The best one i s will therefore be the one with the lowest asymptotic variance V_s when π is fixed.
Using the Central Limit theorem and the Delta Method with $g(u) = (u - 2^{-s})\frac{1}{1-2^{-s}}$, we get :

$$\sqrt{n}(\hat{\pi}_s - \pi) \xrightarrow{D} N(0, V_s)$$

Where $V_s = V(t_s(X))\frac{1}{(1-2^{-s})^2}$
And we have :

$$\begin{aligned} V(t_s(X)) &= \frac{1}{\Gamma(s+1)^2}V(X^s) \\ &= \frac{1}{\Gamma(s+1)^2}(E(X^{2s}) - E(X^s)^2)) \\ &= \frac{1}{\Gamma(s+1)^2}((\pi + (1-\pi)2^{-2s})\Gamma(2s+1) - (\pi + (1-\pi)2^{-s})^2\Gamma(s+1)^2) \end{aligned}$$

Thus: $V_s = \frac{\Gamma(2s+1)}{\Gamma(s+1)^2}(\pi + (1-\pi)2^{-2s})\frac{1}{(1-2^{-s})^2} - (\pi + (1-\pi)2^{-s})^2\frac{1}{(1-2^{-s})^2}$
The right part of the function is increasing with s (this can be shown by deriving the second term). In the left part, the fraction depending on the Gamma function is also inscreasing (using the relationship between Gamma and the factorial function, as s is an integer). The other part of the left part is also an increasing function (this can also be showed by deriving the second term). So V_s is increasing with s. Therefore, we need to choose $s = 1$.

1.16 Bernoulli random variable $Y \sim B((1+e^\theta)^{-1})$

General case : n iid observations $y_1, ..., y_n$
Likelihood : $L(y|\theta) = \prod_{i=1}^n (\frac{1}{1+e^\theta})^{y_i} (\frac{e^\theta}{1+e^\theta})^{1-y_i}$
Log-Likelihood : $l(y|\theta) = \sum_{i=1}^n y_i log(\frac{1}{1+e^\theta}) + (1-y_i) log(\frac{e^\theta}{1+e^\theta})$
Which can be rewritten as : $l(y|\theta) = -nlog(1+e^\theta) + \theta \sum_{i=1}^n (1-y_i)$
Derivative : $\frac{\partial l}{\partial \theta}(y|\theta) = -n\frac{e^\theta}{1+e^\theta} + \sum_{i=1}^n (1-y_i)$
Maximum-Likelihood estimator : it verifies $\frac{\partial l}{\partial \theta}(y|\theta) = 0$
$\frac{1}{n}\sum_{i=1}^n (1-y_i) = 1 - \frac{1}{1+e^\theta}$
$1 - \overline{y} = 1 - \frac{1}{1+e^{\hat{\theta}}}$
$1 + e^{\hat{\theta}} = \frac{1}{\overline{y}}$
$\hat{\theta} = log(\frac{1}{\overline{y}} - 1)$

(a) One observation

If we only have one observation y, then the maximum likelihood estimator is $\hat{\theta} = log(\frac{1}{y} - 1)$. We then see that if $y = 0$, then $\hat{\theta} = +\infty$.

(b) Two observations

If we observe $y_1 = 0$ and $y_2 = 0$, then, like in (a), $\overline{y} = 0$. The maximum likelihood estimator is therefore: $\hat{\theta} = +\infty$.
If we observe $y_1 = 1$ and $y_2 = 1$, then $\overline{y} = 1$, so the maximum likelihood estimator is $\hat{\theta} = log(0) = -\infty$.
Otherwise, if one of the observations is equal to 0 and the other to 1, we get $\overline{y} = 1/2$ so the maximum likelihood estimator is $\hat{\theta} = log(2-1) = 0$.

1.22 Bayes Estimator

(a)

Using the Bayes formula, the Bayes Risk $\int\int L(\delta(x), h(\theta)) f(x|\theta)\pi(\theta) dx d\theta$ is also equal to $\int\int L(\delta(x), h(\theta))\pi(\theta|x)\pi(x) dx d\theta = E_X[E^\pi(L(\delta(X), h(\theta))|X)]$.
If we manage to find $\tilde{\delta}$ such that for each x, $\tilde{\delta}$ minimizes $E^\pi(L(\delta(x), h(\theta))|X = x) = \int L(\delta(x), h(\theta))\pi(\theta|x) d\theta$, then we get:
$\forall \delta, \forall x, E^\pi(L(\tilde{\delta}(x), h(\theta))|X = x) \leq E^\pi(L(\delta(x), h(\theta))|X = x)$
$\forall \delta, E^\pi(L(\tilde{\delta}(X), h(\theta))|X) \leq E^\pi(L(\delta(X), h(\theta))|X)$
So using the Monotonicity of the Expected Value : $\forall \delta, E_X[E^\pi(L(\tilde{\delta}(X), h(\theta))|X)] \leq E_X[E^\pi(L(\delta(X), h(\theta))|X)]$
i.e : $\tilde{\delta}$ minimizes the Bayes Risk.

(b) Mean Squared Error

Case where $L(\delta, h(\theta)) = ||h(\theta) - \delta||^2$

Let x be fixed. Using (a), we try to minimize $E^\pi(||h(\theta) - \delta||^2|x)$.
$E^\pi(||h(\theta) - \delta||^2|x) = V(h(\theta) - \delta(x)|x) + E(h(\theta) - \delta(x)|x)^2$
As $\delta(x)$ is a constant conditionnally on x, we get :
$E^\pi(||h(\theta) - \delta||^2|x) = V^\pi(h(\theta)|x) + [E^\pi(h(\theta)|x) - \delta(x)]^2$
The first part $V^\pi(h(\theta)|x)$ does not depend on δ, so we only need to minimize the second part $[E^\pi(h(\theta)|x) - \delta(x)]^2$ which is minimal when $\delta(x) = E^\pi(h(\theta)|x)$.

(c) Mean Absolute Error

Case where $L(\delta, h(\theta)) = |h(\theta) - \delta|$

Let consider the particular case where $h(\theta) = \theta$.

Let x be fixed. Using (a), we try to minimize $E^\pi(|\theta - \delta||x)$.
$E^\pi(|\theta - \delta||x) = \int |\theta - \delta(x)|\pi(\theta|x) d\theta$

$E^\pi(|\theta-\delta||x) = \int_{\delta(x)}^{+\infty}(\theta-\delta(x))\pi(\theta|x)d\theta + \int_{-\infty}^{\delta(x)}(\delta(x)-\theta)\pi(\theta|x)d\theta$
Let $F(\delta) = E^\pi(|\theta-\delta||x)$.
When we derive F, we get : $F'(\delta) = \int_{-\infty}^{\delta}\pi(\theta|x)d\theta - \int_{\delta}^{+\infty}\pi(\theta|x)d\theta$
So F is minimal when δ verifies : $\int_{-\infty}^{\delta}\pi(\theta|x)d\theta = \int_{\delta}^{+\infty}\pi(\theta|x)d\theta$
As $\int_{-\infty}^{\delta}\pi(\theta|x)d\theta + \int_{\delta}^{+\infty}\pi(\theta|x)d\theta = 1$, we get :
$2\int_{-\infty}^{\delta}\pi(\theta|x)d\theta = 1$, which means that δ is the median of the posterior distribution $\pi(\theta|x)$.

In the more general case where h is monotonic and C^1, we need to use first Integration by Substitution with $h(\theta) = u$ in order to remove the absolute value. We would finally get an estimator equal to the median of the posterior distribution of $h(\theta)$ (and not the median of the posterior distribution of θ).