Fundamentals of Inference and Learning

Homework 1

1 Statistical Inference & Maximum Likelihood

$$p_\lambda(x_i) = \begin{cases} \frac{1}{Z(\lambda)} e^{-x_i/\lambda} & \text{if } 1 \leq x_i \leq 20 \\ 0 & \text{otherwise} \end{cases}$$

λ^* : true λ

is in the range $[0.05 ; 50]$

1/ $$\int_{-\infty}^{+\infty} p_\lambda(x_i)\, dx_i = 1$$

$$\Leftrightarrow \int_1^{20} \frac{1}{Z(\lambda)} e^{-x_i/\lambda}\, dx_i =$$

$$-\frac{\lambda}{Z(\lambda)} e^{-x_i/\lambda} \Big|_1^{20} = -\frac{\lambda}{Z(\lambda)}\left(e^{-20/\lambda} - e^{-1/\lambda}\right) = 1$$

$$\Leftrightarrow Z(\lambda) = -\lambda\left(e^{-20/\lambda} - e^{-1/\lambda}\right)$$

$$\mathbb{E}\{X\} = \int_{-\infty}^{+\infty} x\, p_\lambda(x)\, dx = \int_1^{20} \frac{x}{Z(\lambda)} e^{-x/\lambda}\, dx$$

$-\lambda e^{-x/\lambda}$	$e^{-x/\lambda}$
x	1

$$= \frac{1}{Z(\lambda)} \left\{ \left[-\lambda e^{-x/\lambda} x\right]_1^{20} + \lambda \int_1^{20} e^{-x/\lambda}\, dx \right\}$$

$$= \frac{1}{Z(\lambda)} \left\{ \left[-\lambda e^{-x/\lambda} x\right]_1^{20} - \lambda^2 \left[e^{-x/\lambda}\right]_1^{20} \right\}$$

$$= \frac{1}{Z(\lambda)} \left\{ -\lambda^2\left(e^{-20/\lambda} - e^{-1/\lambda}\right) - \lambda\left(20e^{-20/\lambda} - e^{-1/\lambda}\right) \right\}$$

$$= \frac{-\lambda}{Z(\lambda)} \left[\lambda\left(e^{-20/\lambda} - e^{-1/\lambda}\right) + \left(20e^{-20/\lambda} - e^{-1/\lambda}\right) \right]$$

$$= \frac{\lambda}{Z(\lambda)} \left[Z(\lambda) - \left(20e^{-20/\lambda} - e^{-1/\lambda}\right) \right]$$

$$\partial_\lambda \log Z(\lambda) = \partial_\lambda \log(-\lambda(e^{-20/\lambda} - e^{-1/\lambda}))$$

$$= \frac{-1}{\lambda(e^{-20/\lambda}-e^{-1/\lambda})}\left[-(e^{-20/\lambda}-e^{-1/\lambda}) - \lambda\left(\frac{-20e^{-20/\lambda}}{\lambda^2} - \frac{-e^{-1/\lambda}}{\lambda^2}\right)\right]$$

$$= \frac{1}{Z(\lambda)}\left[-(e^{-20/\lambda}-e^{-1/\lambda}) + \frac{20e^{-20/\lambda}}{\lambda} - \frac{e^{-1/\lambda}}{\lambda}\right]$$

By multiplying by λ^2

$$= \frac{1}{Z(\lambda)}\left[\lambda^2(e^{-20/\lambda}-e^{-1/\lambda}) - \lambda(20e^{-20/\lambda} - e^{-1/\lambda})\right]$$

$$= \frac{\lambda}{Z(\lambda)}\left[Z(\lambda) - (20e^{-20/\lambda} - e^{-1/\lambda})\right]$$

Conversely one can compute $Z'(\lambda)$ & $Z''(\lambda)$:

- $Z(\lambda) = \lambda(e^{-1/\lambda} - e^{-20/\lambda}) = \lambda[e^{-x/\lambda}]_{20}^{1}$
- $Z'(\lambda) = (e^{-1/\lambda}-e^{-20/\lambda}) + \cancel{\lambda}\left(\frac{e^{-1/\lambda}}{\lambda^{\cancel{2}}} - 20\frac{e^{-20/\lambda}}{\lambda^{\cancel{2}}}\right)$
- $\lambda Z'(\lambda) = Z(\lambda) + [xe^{-x/\lambda}]_{20}^{1}$
- $\frac{\partial}{\partial\lambda}\lambda Z'(\lambda) = \cancel{Z'(\lambda)} + \lambda Z''(\lambda) = \cancel{Z'(\lambda)} + \left(\frac{e^{-1/\lambda}}{\lambda^2} - 400\frac{e^{-20/\lambda}}{\lambda^2}\right)$

  $\Leftrightarrow \lambda Z''(\lambda) = \left(\frac{e^{-1/\lambda}}{\lambda^2} - 400\frac{e^{-20/\lambda}}{\lambda^2}\right)$
- $\lambda^4 Z''(\lambda) = \lambda[x^2 e^{-x/\lambda}]_{20}^{1}$

Now let's compute the variance:

$$\mathrm{Var}[X] = \underbrace{\mathbb{E}[X^2]}_{1)} - \underbrace{(\mathbb{E}[X])^2}_{2)}$$

1) $\mathbb{E}[X]^2 = \left(\lambda^2 \frac{Z'(\lambda)}{Z(\lambda)}\right)^2 = \lambda^4 \cdot \frac{Z'(\lambda)\cdot Z'(\lambda)}{Z(\lambda)^2}$

2) $\mathbb{E}[X^2] = \frac{1}{Z(\lambda)} \int_1^{20} x^2 e^{-x/\lambda} dx = \frac{1}{Z(\lambda)} \left(\underbrace{[-x^2\lambda e^{-x/\lambda}]_1^{20}}_{= \lambda^4 Z''(\lambda)} + 2\lambda \underbrace{\int_1^{20} x e^{-x/\lambda} dx}_{\mathbb{E}[\lambda]} \right)$

$$= \frac{1}{Z(\lambda)} \left(\lambda^4 Z''(\lambda) + 2\lambda \mathbb{E}[\lambda] \right)$$

$$= \frac{\lambda^4 Z''(\lambda)}{Z(\lambda)} + 2\lambda^3 \frac{Z'(\lambda)}{Z(\lambda)}$$

$$\Rightarrow \operatorname{Var}[X] = \frac{\lambda^4 Z''(\lambda)}{Z(\lambda)} + 2\lambda^3 \frac{Z'(\lambda)}{Z(\lambda)} - \lambda^4 \cdot \frac{Z'(\lambda) \cdot Z'(\lambda)}{Z(\lambda)^2}$$

$$= 2\lambda^3 \frac{Z'(\lambda)}{Z(\lambda)} + \lambda^4 \frac{Z''(\lambda)Z(\lambda) - Z'(\lambda)Z'(\lambda)}{Z(\lambda)^2}$$

$$= \frac{\partial}{\partial\lambda} \left(\lambda^4 \cdot \frac{Z'(\lambda)}{Z(\lambda)} \right) = \frac{\partial}{\partial\lambda} \left(\lambda^2 \mathbb{E}[X] \right) \quad \square$$

2) $p_\lambda(\{x_i\}_{i=1}^n) = \prod_{i=1}^n p_\lambda(x_i)$

$$\log p_\lambda(\{x_i\}_{i=1}^n) = \log \prod_{i=1}^n p_\lambda(x_i) = \sum_{i=1}^n \log p_\lambda(x_i) = \sum_{i=1}^n \log \left(\frac{1}{Z(\lambda)} e^{-x_i/\lambda} \right)$$

$$= \sum_{i=1}^n -\frac{x_i}{\lambda} - \log Z(\lambda) = \sum_{i=1}^n -\frac{x_i}{\lambda} - \log[-\lambda(e^{-20/\lambda} - e^{-1/\lambda})]$$

$$= -n \log[-\lambda(e^{-20/\lambda} - e^{-1/\lambda})] - \frac{1}{\lambda} \sum_{i=1}^n x_i$$

$$\mathcal{L}_\lambda(\{x_i\}_{i=1}^n) = -\log[-\lambda(e^{-20/\lambda} - e^{-1/\lambda})] - \frac{1}{n\lambda} \sum_{i=1}^n x_i$$

3) see Notebook

4) · $\hat{\lambda}_{ML}(\{x_i\}_{i=1}^n) = \operatorname{argmax}_\lambda \log(p_\lambda(\{x_i\}_{i=1}^n))$

· $SE = (\hat{\lambda}_{ML}(\{x\}) - \lambda^*)^2$

5) $\hat{\lambda}_J(\{x_i\}_{i=1}^{N}) = \underset{\lambda}{\operatorname{argmax}} \log(p_\lambda(\{x_i\}_{i=1}^{n})\sqrt{I(\lambda)}) = \underset{\lambda}{\operatorname{argmax}} \log(p_\lambda(\{x_i\}_{i=1}^{n})) + \log(\sqrt{I(\lambda)})$

$= \operatorname{argmax} \log(p_\lambda(\{x_i\}_{i=1}^{n})) + \log\left(\frac{\sqrt{Var[X]}}{\lambda^2}\right)$

2 Probability bounds & a pooling problem

1/ $P(\frac{1}{m}\sum_i z_i \geq p+\varepsilon) \leq e^{-2m\varepsilon^2}$

$\Leftrightarrow P(\sum_i z_i \geq m(p+\varepsilon)) \leq e^{-2m\varepsilon^2}$

using the hint given in the problem set:

$P(e^{\lambda \sum z_i} \geq e^{\lambda m(p+\varepsilon)}) \leq e^{-2m\varepsilon^2}$

Now let's apply Markov's inequality

$$P(e^{\lambda \sum z_i} \geq e^{\lambda m(p+\varepsilon)}) \leq \frac{\mathbb{E}[e^{\lambda \sum_i z_i}]}{e^{\lambda m(p+\varepsilon)}} = \frac{\prod_i \mathbb{E}[e^{\lambda z_i}]}{(e^{\lambda(p+\varepsilon)})^m}$$

· for a Bernoulli variable, $\mathbb{E}[e^{\lambda z_i}] = \sum_i e^{\lambda z_i} P(Z=z_i)$

$$= e^0 \cdot (1-p) + e^\lambda p = pe^\lambda + (1-p)$$

$$\Rightarrow P(e^{\lambda \sum z_i} \geq e^{\lambda m(p+\varepsilon)}) \leq \frac{(pe^\lambda + (1-p))^m}{(e^{\lambda(p+\varepsilon)})^m} = \left(\frac{pe^\lambda + (1-p)}{e^{\lambda(p+\varepsilon)}}\right)^m \quad \square$$

2/ $P(\frac{1}{m}\sum_i z_i \geq p+\varepsilon) \leq e^{-mf(p,\varepsilon)}$

$$f(p,\varepsilon) = (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) + (1-(p+\varepsilon))\log\left(\frac{1-(p+\varepsilon)}{1-p}\right)$$

we have to differentiate

$$\log\left(\left(\frac{pe^\lambda + (1-p)}{e^{\lambda(p+\varepsilon)}}\right)^m\right) = m\log\left(\frac{pe^\lambda + (1-p)}{e^{\lambda(p+\varepsilon)}}\right) \quad \text{over } \lambda$$

$$\frac{\partial}{\partial\lambda}\log\left(\left(\frac{pe^\lambda + (1-p)}{e^{\lambda(p+\varepsilon)}}\right)^m\right) = \frac{\partial}{\partial\lambda}\left[m(\log(pe^\lambda + (1-p)) - \log(e^{\lambda(p+\varepsilon)})\right]$$

$$= m\frac{\partial}{\partial\lambda}(\log(pe^\lambda + (1-p)) - \log(e^{\lambda(p+\varepsilon)})$$

$$= m\left[\frac{pe^\lambda}{pe^\lambda + (1-p)} - \frac{(p+\varepsilon)\cancel{e^{\lambda(p+\varepsilon)}}}{\cancel{e^{\lambda(p+\varepsilon)}}}\right] = 0$$

$$\Leftrightarrow \frac{pe^\lambda}{pe^\lambda + (1-p)} = p+\varepsilon \Leftrightarrow pe^\lambda = (p+\varepsilon)(pe^\lambda + (1-p))$$

$$\Leftrightarrow pe^\lambda(1-(p+\varepsilon)) = (1-p)(p+\varepsilon) \Leftrightarrow e^\lambda = \frac{(1-p)(p+\varepsilon)}{p(1-(p+\varepsilon))}$$

$$\left(\frac{pe^{\lambda}+(1-p)}{e^{\lambda(p+\varepsilon)}}\right)^{m} = e^{-mf(p,\varepsilon)}$$

$$\Leftrightarrow \log\left(\frac{pe^{\lambda}+(1-p)}{e^{\lambda(p+\varepsilon)}}\right)^{\cancel{m}} = -\cancel{m}f(p,\varepsilon)$$

$$f(p,\varepsilon) = -\log(pe^{\lambda}+(1-p)) + \log(e^{\lambda(p+\varepsilon)})$$

$$= \underbrace{-\log\left(\cancel{p}\frac{(1-p)(p+\varepsilon)}{\cancel{p}(1-(p+\varepsilon))}+(1-p)\right)}_{2)} + \underbrace{(p+\varepsilon)\log\left(\frac{(1-p)(p+\varepsilon)}{p(1-(p+\varepsilon))}\right)}_{1)}$$

1) $(p+\varepsilon)\left[\log(1-p)+\log\left(\frac{p+\varepsilon}{p}\right)-\log(1-(p+\varepsilon))\right]$

$$= (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) - (p+\varepsilon)\log\left(\frac{1-p+\varepsilon}{1-p}\right)$$

2) $-\log\left((1-p)\underbrace{\left(\frac{p+\varepsilon}{1-(p+\varepsilon)}+1\right)}_{3)}\right) = -\log\left(\frac{1-p}{1-p+\varepsilon}\right) = \log\left(\frac{1-p+\varepsilon}{1-p}\right)$

3) $\frac{1-(p+\varepsilon)+(p+\varepsilon)}{1-p+\varepsilon} = \frac{1}{1-p+\varepsilon}$

By combining 1) & 2)

$$f(p,\varepsilon) = (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) - (p+\varepsilon)\log\left(\frac{1-p+\varepsilon}{1-p}\right) + \log\left(\frac{1-p+\varepsilon}{1-p}\right)$$

$$= (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) + (1-(p+\varepsilon))\log\left(\frac{1-p+\varepsilon}{1-p}\right)$$

Finally:

$$P\left(\frac{1}{m}\sum_i z_i \geq p+\varepsilon\right) \leq e^{-mf(p,\varepsilon)}$$

$$f(p,\varepsilon) = (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) + (1-(p+\varepsilon))\log\left(\frac{1-(p+\varepsilon)}{1-p}\right) \quad \square$$

3) $f(p,\varepsilon=0)= p\log(p/p)+(1-p)\log\left(\frac{1-p}{1-p}\right)=0$

$$\frac{\partial f(p,\varepsilon)}{\partial\varepsilon}\Big|_{\varepsilon=0} = \log\left(\frac{p+\varepsilon}{p}\right)+(p+\varepsilon)\cdot\frac{1}{p+\varepsilon} - \log\left(\frac{1-(p+\varepsilon)}{1-p}\right) - (1-(p+\varepsilon))\frac{1}{1-(p+\varepsilon)}\Big|_{\varepsilon=0}$$

$$= \log(p/p)+1-\log\left(\frac{1-p}{1-p}\right)-1 = 0+1-0-1=0$$

$$\frac{\partial^2 f(p,\varepsilon)}{\partial\varepsilon^2} = \frac{\partial}{\partial\varepsilon}\left(\frac{\partial f(p,\varepsilon)}{\partial\varepsilon}\right) = \frac{1}{p+\varepsilon}+\frac{1}{1-(p+\varepsilon)} = \frac{1}{(p+\varepsilon)(1-(p+\varepsilon))} = \underbrace{\frac{1}{-(p+\varepsilon)^2+(p+\varepsilon)}}_{1)} \geq 4$$

1) $h(x)\frac{1}{-x^2+x} \geq 4$

$h'(x) = \frac{-2x+1}{(-x^2+x)^2} = 0 \iff x = 1/2$

$h''(x) = \frac{2(1-2x)^2}{(x-x^2)^3} + \frac{2}{(x-x^2)^2}$

	0		1/2		1
$f'(x)$		↘	0	↗	
$f(x)$		↘	4	↗	
$f''(x)$		↘	0	↗	

$\Rightarrow x=\frac{1}{2}$ is a global minimum. Therefore $h(x)\geq 4$

4) $f(p,\varepsilon) = f(p,0)+\varepsilon f'(p,0)+\varepsilon^2\frac{f''(p,\tilde{\varepsilon})}{2} = \varepsilon^2\,\frac{f''(p,\tilde{\varepsilon})}{2}$

$\iff f(p,\varepsilon) = \frac{f''(p,\tilde{\varepsilon})}{2}\varepsilon^2$

Then using the previous result $f''(p,\tilde{\varepsilon})\geq 4$

$f(p,\varepsilon)\geq 2\varepsilon^2$

Combining this result with $\left(\frac{pe^{\lambda}+(1-p)}{e^{\lambda(p+\varepsilon)}}\right)^m = e^{-mf(p,\varepsilon)}$

leads to

$\left(\frac{pe^{\lambda}+(1-p)}{e^{\lambda(p+\varepsilon)}}\right)^m \leq e^{-2m\varepsilon^2}$

And finally $P(e^{\lambda\Sigma z_i}\geq e^{\lambda m(p+\varepsilon)}) \leq \left(\frac{pe^{\lambda}+(1-p)}{e^{\lambda(p+\varepsilon)}}\right)^m \leq e^{-2m\varepsilon^2}$

Bonus 2

$$P\left(\left|\frac{1}{m}\sum_i z_i - p\right| \geq \varepsilon\right) = P\left(\frac{1}{m}\sum_i z_i - p \geq \varepsilon \,,\, \frac{1}{m}\sum_i z_i - p \leq \varepsilon\right)$$

$$= P\left(\frac{1}{m}\sum_i z_i \geq \varepsilon + p\right) + P\left(\frac{1}{m}\sum_i z_i \geq \varepsilon + p\right) \leq 2e^{-2m\varepsilon^2}$$

5/ Using $P\left(\left|\frac{1}{m}\sum_i z_i - p\right| \geq \varepsilon\right) \leq 2e^{-2m\varepsilon^2}$

with $\varepsilon = 0.01$ & $2e^{-2m\varepsilon^2} = 0.05$ as we want that the probability to be outside the confidence interval to be smaller than 5%

$$P\left(\left|\frac{1}{m}\sum_i z_i - p\right| \geq 0.01\right) \leq 0.05$$

which leads to $2e^{-2m\varepsilon^2} = 0.05$

$$m = -\frac{\log(0.05/2)}{2 \cdot 0{,}01^2} = 18\,445 \text{ people}$$

6/ see notebook