SB2.1 Foundations of Statistical Inference

Sheet 1 — MT22

For Tutors Only — Not For Distribution

Section A

1. Let $X_1, \ldots, X_n$ be independent Poisson random variables with means $\mathbb{E}(X_i) = \lambda m_i$, $i = 1, \ldots, n$ where $\lambda > 0$ is unknown and $m_1, \ldots, m_n$ are known constants.

   (a) Show that the model defines an exponential family with canonical parameter $\theta = \log \lambda$.

   (b) What is the canonical observation? Find its mean and variance.

   (c) Find the MLE $\widehat{\theta}$ of θ.

   (d) What can we say about $\mathbb{E}[\widehat{\theta}]$?

   (e) Show that for any function $T : \mathbb{N} \mapsto \mathbb{R}$ we have that

$$\lim_{\lambda \to 0} \mathbb{E}_\lambda \left[T\Big(\sum_{i=1}^n X_i \Big) \right] = T(0).$$

   (f) Conclude that there cannot exist an unbiased estimator of θ.

Solution:

(a)

$$\begin{aligned} L(\lambda, \mathbf{x}) &= \prod_1^n e^{-\lambda m_i} (\lambda m_i)^{x_i} / x_i \\ &= \exp\left\{ (\log \lambda) \sum_1^n x_i - \lambda \sum_1^n m_i + \sum_1^n x_i \log m_i - \sum_1^n \log(x_i!) \right\} \end{aligned}$$

which is in canonical exponential form with $\theta = \log \lambda$, $B_1(x) = \sum_1^n x_i$

(b) The canonical (minimal) sufficient statistic is $\bar{X}$ ($n\bar{X}$ is fine as well). $\mathbb{E}[\bar{X}] = \lambda \bar{m}$. $\sum_1^n X_i$ is Poisson $(\lambda \sum_1^n m_i)$ so $\mathrm{Var}(\bar{X}) = \lambda \bar{m}/n$.

(c) $\ell(\theta) = \text{const} + \theta \sum_1^n x_i - e^\theta \sum_1^n m_i$, $\partial \ell / \partial \theta = \sum_1^n x_i - e^\theta \sum_1^n m_i$, so $\widehat{\theta} = \log[\bar{x}/\bar{m}]$, provided $\bar{x} > 0$. If $\bar{x} = 0$ then $\widehat{\lambda} = 0$, $\widehat{\theta} = -\infty$. As $n \to \infty$ the probability that $\bar{X} = 0$ tends to zero if $\lambda > 0$.

(d) In fact, $E[\widehat{\theta}]$ does not even exist. This is because $P(\widehat{\theta} = -\infty) = P(\bar{x} = 0) > 0$.

George Deligiannidis: deligian@stats.ox.ac.uk

(e) $\sum X_i \sim \text{Poi}(\lambda \sum m_i)$ Without loss of generality assume that $\sum m_i = 1$ to simplify notations. So we want to prove that if $X \sim \text{Poi}(\lambda)$ under $\mathbb{E}_\lambda$ then

$$\lim_{\lambda \to 0} \mathbb{E}_\lambda[T(X)] = T(0)$$

Observe that

$$E_\lambda[T] = T(0)\mathrm{e}^{-\lambda} + \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} T(k)\frac{\lambda^k}{k!}.$$

Notice that if $\lambda_1 < \lambda_2$ then for $k \geq 1$ we have $\lambda_1^k < \lambda_2^k$ and thus we also have

$$|T(k)|\frac{\lambda_1^k}{k!} \leq |T(k)|\frac{\lambda_2^k}{k!}.$$

Now we need to assume that T is integrable at least for one $\lambda_0 < 1$ so that

$$\mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{|T(k)|\lambda^k}{k!} < \infty.$$

Therefore we have that for $\lambda < \lambda_0$

$$\frac{T(k)\lambda^k}{k!}$$

are dominated by the summable $T(k)\lambda_0^k/k!$, and for each $k \geq 1$, $T(k)\lambda^k/k! \to 0$. We apply the dominated convergence theorem to obtain that

$$\lim_{\lambda \to 0} \sum_{k \geq 1} T(k)\frac{\lambda^k}{k!} = 0,$$

and therefore, since $\mathrm{e}^{-\lambda} \to 1$ as $\lambda \to 0$, we also have that

$$\lim_{\lambda \to 0} \mathrm{e}^{-\lambda} \sum_{k \geq 1} T(k)\frac{\lambda^k}{k!} = \lim_{\lambda \to 0} \sum_{k \geq 1} T(k)\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} = 0.$$

Thus

$$\mathbb{E}_\lambda[T] \to T(0) \qquad \text{as } \lambda \to 0$$

(f) Notice that since $\theta = \log(\lambda)$, we have that $\theta \to -\infty$ as $\lambda \to 0$. Therefore if T is any unbiased estimator, then for any $K > 0$ we can find $\epsilon > 0$ such that for $\lambda < \epsilon$ $\mathbb{E}_\lambda[T] < -K$. But $T(0) > -\infty$ and therefore we arrive at a contradiction.

Department of Statistics, University of Oxford
George Deligiannidis: deligian@stats.ox.ac.uk

Section B

2. Let $X_1, \ldots, X_n$ be a random sample from the density

$$f(x; \theta) = e^{-(x-\theta)},\ x \geq \theta$$

   (a) Show that the MLE $\widehat{\theta}$ of θ is the minimum of $X_1, \ldots, X_n$.

   (b) Show that $\hat{\theta}$ is a sufficient for θ.

   (c) Show that for all $\epsilon > 0$

$$P_\theta[|\widehat{\theta} - \theta| > \epsilon] \leq e^{-n\epsilon},$$

   deduce that $\widehat{\theta}$ is consistent in probability and in quadratic mean, that is $\widehat{\theta} \to \theta$ in probability and in L^2 (we say that $X_n \to X$ in L^2 if $E[(X_n - X)^2] \to 0$), but that it is a biased estimator of θ with $\mathbb{E}[\widehat{\theta}] = \theta + 1/n$. Suggest an unbiased and consistent estimator and find its variance.

Solution:

$$L(\theta; \mathbf{x}) = e^{-\sum_1^n x_i + n\theta} \prod_{i=1}^n \mathbb{I}_{[x_i > \theta]} = e^{n\theta} e^{-\sum x_i} \mathbb{I}_{[\min x_i \geq \theta]}$$

Note that X_1 is just θ plus a mean 1 exponential r.v.

(a) To maximize $L(\theta, x)$ we need to look at the boundaries. Once we do that it is clear that $\widehat{\theta} = \min_i X_i$ maximizes $L(\theta, \mathbf{x})$.

(b)

$$L(\theta; \mathbf{x}) = e^{n\theta} \mathbb{I}_{\min(x_i) > \theta} \times e^{-n\bar{x}}$$

so it factorizes into $f_1(\min(x_i); \theta) h(x)$ and $\min(x_i)$ is sufficient. The family is not an exponential family

(c) Writing $Z_i = X_i - \theta$, the Z_i are iid Exp(1) r.v. Thus $\hat{\theta} = \theta + \min Z_i$. Remember that $\min_{i=1,\ldots,n} Z_i$ is itself an Exp(n) r.v.

$$P(\widehat{\theta} - \theta > \epsilon) = P(\min Z_i > \epsilon) = e^{-n\epsilon},\ \epsilon > 0$$

and

$$P(\widehat{\theta} - \theta < -\epsilon) = 0$$

Hence for all $\epsilon > 0$

$$\lim_n P_\theta[|\widehat{\theta} - \theta| > \epsilon] = 0$$

George Deligiannidis: deligian@stats.ox.ac.uk

and it is consistent in probability. We also have that

$$E[(\hat{\theta}-\theta)^2] = E[(\min Z_i)^2] = 2/n^2 \to 0$$

as $n \to \infty$ so it is consistent in quadratic mean.

Finally

$$E[\hat{\theta}] = \theta + 1/n,\ V(\hat{\theta}) = n^{-2}.$$

It is not unbiased . An unbiased estimator is $\tilde{\theta} = \hat{\theta} - 1/n$ and its variance is the same as that of $\hat{\theta}$.

3. Let $X = (X_1, \ldots, X_n)$ be an i.i.d. sample from a distribution with density

$$f(x;\theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x},\ x > 0.$$

   (a) Rewrite the density in standard exponential form.

   (b) Find a minimal sufficient statistic for θ, $T(X)$. Find the expected value of the statistic.

   (c) Find the maximum likelihood estimator for θ. Is it unbiased for θ?

   (d) Show that $\theta^* = (2/n)\sum_{i=1}^n X_i^{-1}$ is an unbiased estimator of θ and find its variance.

   (e) Compute the Fisher information $I_n(\theta)$ of the model and compare the variance of θ^* with $I_n(\theta)$.

   [Hint: Recall: The Gamma density with parameters (α,β) is $\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$. If $X \sim \Gamma(a_1,\beta), Y \sim \Gamma(a_2,\beta)$ and independent then $X+Y \sim \Gamma(a_1+a_2,\beta)$. Mean of $\Gamma(\alpha,\beta)$ is α/β.]

Solution:

(a)

$$f(x;\theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x} = \exp\{-\theta x + 3\log\theta\}\,\frac{x^2}{2}$$

is in the standard form with $T(x) = x$, $\eta(\theta) = \theta$, $B(\theta) = 3\log\theta$ and $h(x) = x^2/2$. It is clear that this is a strictly 1-parameter exponential family.

(b)

$$L(\theta;x) \propto \theta^{3n} e^{-\theta\sum x_i} \times \prod x_i^2$$

By the factorization theorem (or by standard results about exponential family) $\bar{x}$ is a minimal sufficient statistic for θ . From the hint we can see that $f(x;\theta)$ is a $\Gamma(3,\theta)$ family and therefore the mean is $3/\theta$.

George Deligiannidis: deligian@stats.ox.ac.uk

(c) $l(\theta) = 3n \log\theta - \theta \sum x_i + \text{const}$, $l'(\theta) = 3n/\theta - \sum x_i$ so $\hat{\theta} = 3/\bar{x}$.

Recall: The Gamma density with parameters (α, β) is $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$. If $X \sim \Gamma(a_1, \beta), Y \sim \Gamma(a_2, \beta)$ and independent then $X + Y \sim \Gamma(a_1 + a_2, \beta)$. Mean of $\Gamma(\alpha, \beta)$ is α/β.

$\sum_1^n X_i$ has a Gamma distribution with density

$$\frac{\theta^{3n}}{\Gamma(3n)} x^{3n-1} e^{-\theta x} \; x > 0$$

so

$$\begin{aligned}
\mathbb{E}[\widehat{\theta}] &= 3n \int_0^\infty \frac{\theta^{3n}}{\Gamma(3n)} x^{3n-2} e^{-\theta x} dx \\
&= 3n \cdot \frac{\theta^{3n}}{\Gamma(3n)} \cdot \frac{\Gamma(3n-1)}{\theta^{3n-1}} \\
&= \frac{3n\theta}{3n-1}
\end{aligned}$$

Thus $\widehat{\theta}$ is a biased estimate of θ.

(d)

$$\begin{aligned}
\mathbb{E}[X_i^{-1}] &= \int_0^\infty \frac{1}{2} \theta^3 x e^{-\theta x} dx \\
&= \frac{1}{2}\theta
\end{aligned}$$

so $\theta^* = (2/n) \sum_1^n X_i^{-1}$ is an unbiased estimate of θ. Similarly, from the density, one can show that $\mathrm{Var}(\theta^*) = \theta^2/n$.

(e) Fisher's information is $I_n(\theta) = -E[\frac{\partial^2}{\partial\theta^2} \ell(\theta)] = 3n/\theta^2$. To find the variance we compute

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{X_i}\right) &= \mathbb{E}\left[X_i^{-2}\right] - \mathbb{E}\left[X_i^{-1}\right]^2 \\
&= \int \frac{1}{2}\theta^3 \mathrm{e}^{-\theta x} x - \left(\frac{\theta}{2}\right)^2 \\
&= \frac{\theta^2}{2} - \frac{\theta^2}{4} = \frac{\theta^2}{4}.
\end{aligned}$$

So $\mathrm{Var}(\theta^*) = \theta^2/n \geq I_n(\theta) = \theta^2/(3n)$.

George Deligiannidis: deligian@stats.ox.ac.uk

4. Let $X_1, \ldots, X_n$ be a sample from $N(\mu, \sigma^2)$.

   (a) Show that the MLE of σ^2 is

$$\widehat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

   (b) Show that $\widehat{\sigma}^2$ has a smaller mean square error than

$$(n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

   (c) For which value of a is the MSE of

$$(n+a)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

   the smallest.

   Hint: For (b) and (c) you will need to find $\mathrm{Var}(\chi^2_{n-1})$ which is a special case of the variance of a gamma distribution.

Solution:

(a)

$$\ell(\mu, \sigma^2) = \text{const} - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{1}^{n} (x_i - \mu)^2 / \sigma^2$$

so

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{1}^{n} (x_i - \mu)^2 \tag{1}$$

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum (x_i - \mu). \tag{2}$$

Setting both equal to 0 we get that $\mu_{\text{MLE}} = \bar{x}$, uniformly in σ^2, so

$$\widehat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Technically we should also do a second derivative test to verify it's indeed a maximum. Recap from Part A Statistics

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\widehat{\sigma}^2 = \frac{n-1}{n} S^2 \sim \frac{\sigma^2}{n} \chi^2_{n-1}$$

George Deligiannidis: deligian@stats.ox.ac.uk

χ^2_r has a density

$$\frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, \quad > x > 0$$

which is a $\Gamma(r/2, 1/2)$ density with mean $2 \times r/2 = r$ and variance $4 \times r/2 = 2r$.

(b) $\mathbb{E}(\widehat{\sigma}^2) = ((n-1)/n)\sigma^2$, $\text{Bias}(\widehat{\sigma}^2) = -\sigma^2/n$, $\text{Var}(\widehat{\sigma}^2) = (2(n-1)/n^2)\sigma^4$. Thus

$$\text{MSE}(\widehat{\sigma}^2) = \text{Var}(\widehat{\sigma}^2) + \text{Bias}(\widehat{\sigma}^2)^2 = \frac{2n-1}{n^2}\sigma^4$$

Let

$$S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \frac{\sigma^2}{n-1} \chi^2_{n-1},$$

then $[S^2] = \sigma^2$, so unbiased. Therefore the MSE is simply the variance and therefore

$$\text{MSE}(S^2) = \frac{2(n-1)}{(n-1)^2}\sigma^4 = \frac{2}{n-1}\sigma^4 > \text{MSE}(\widehat{\sigma}^2) = \frac{2n-1}{n^2}\sigma^4.$$

(c) Let

$$\sigma^{*2} = (n+a)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

A similar calculation to (b) shows that

$$\text{MSE}(\sigma^{*2}) = \left(\frac{2}{n-1} b^2 + (b-1)^2 \right) \sigma^4$$

where $b = (n-1)/(n+a)$. The MSE is minimal when

$$b = \frac{1}{\frac{2}{n-1} + 1}, \text{ or } a = 1$$

That is the minimal MSE solution is

$$(n+1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

George Deligiannidis: deligian@stats.ox.ac.uk

5. (a) Let $Y_1, \dots, Y_n$ be a random sample from a Poisson distribution with parameter $\lambda > 0$. One observes only $W_i = \mathbf{1}_{Y_i > 0}$. Compute the likelihood associated with the sample $(W_1, \dots, W_n)$ and the MLE in λ. Show that it is consistent in probability.

   (b) Let $X_1, \dots, X_n$ be a random sample from a truncated Poisson distribution with distribution
$$f(x; \lambda) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \cdot \frac{\lambda^x}{x!}, \; x = 1, 2, \dots.$$
For $i = 1, \dots, n$ a random variable Z_i is defined by
$$Z_i = X_i \text{ if } X_i \geq 2 \text{ or } Z_i = 0 \text{ if } X_i = 1$$
Show that $\bar{Z}$ is an unbiased estimator of λ with efficiency (efficiency is the ratio of the variance to the Cramer-Rao lower bound)
$$\frac{1 - e^{-\lambda}}{1 - \left(\frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}}\right)^2}.$$

Solution:

(a) For the first part, the likelihood function is the folowing: let $w = (w_1, \dots, w_n)$ be the vector of observations and let $S = \sum w_i$. Then
$$L(\lambda, w) = (1 - e^{-\lambda})^S e^{-\lambda(n-S)} = (e^{\lambda} - 1)^S e^{-n\lambda}$$
So that
$$\ell'(\lambda) = S\frac{e^{\lambda}}{e^{\lambda} - 1} - n$$
and solving $\ell' = 0$ gives us
$$\hat{\lambda} = -\log(1 - \frac{S}{n}).$$
Note that we have again the problem that $\hat{\lambda} = \infty$ with positive probability.

Observe that the W_i are iid Bernoulli variables with parameter $p = 1 - e^{-\lambda}$. The MLE estimator for p is well known to be $\hat{p} = S/n$. Notice that $\hat{p} = 1 - e^{-\hat{\lambda}}$ (or $\hat{\lambda} = -\log(1 - \hat{p})$). This is an example of the invariance of the MLE w.r.t. one-to-one reparametrization.

Notice that $p \mapsto \log(1 - p)$ is uniformly continuous on $[0, 1 - \delta]$ for any $\delta > 0$. Suppose first that $\lambda < \infty$, or equivalently that $p = 1 - \mathrm{e}^{-\lambda} < 1 - \delta$ for some $\delta > 0$. Then there exists a $K = K_\delta$ such that
$$|\log(1 - p) - \log(1 - p')| \leq K_\delta |p - p'|, \quad \text{for all } p, p' \in [0, 1 - \delta].$$

George Deligiannidis: deligian@stats.ox.ac.uk

Then we have for any $\epsilon > 0$

$$\begin{aligned}\mathbb{P}\left[|\log(1-\hat{p}_n) - \log(1-p)| > \epsilon\right] &\le \mathbb{P}\Big[\big\{|\log(1-\hat{p}_n) - \log(1-p)| > \epsilon\big\} \cap \big\{|\hat{p}_n - p| \le \delta/2\big\}\Big] \\ &\quad + \mathbb{P}\left[|\hat{p}_n - p| > \delta/2\right] \\ &\le \mathbb{P}\Big[\,|\hat{p}_n - p| > \epsilon/K_{\delta/2}\Big] + o(1) = o(1)\end{aligned}$$

by consistency of $\hat{p}_n$.

On the other hand if $\lambda = \infty$, then $p = 1$ we have that $S/n = \hat{p} = 1 = p$ with probability 1. Therefore $\log(1-\hat{p}) = +\infty = \lambda$ with probability 1. Therefore we have consistency.

(b) For the second part,

$$f(x;\lambda) = \frac{e^{-\lambda}}{1-e^{-\lambda}}\frac{\lambda^x}{x!}, x = 1, 2, \dots$$

The mean of Z is

$$\mathbb{E}[Z] = \sum_{x\ge 2} x\frac{e^{-\lambda}}{1-e^{-\lambda}}\frac{\lambda^x}{x!} = \frac{e^{-\lambda}}{1-e^{-\lambda}}\sum_{x\ge 2}\frac{\lambda^x}{(x-1)!} = \frac{e^{-\lambda}}{1-e^{-\lambda}}\lambda(e^{\lambda}-1) = \lambda$$

Therefore $\bar{Z} = \sum Z_i/n$ is an unbiased estimator.

Now we want to compute the efficiency. For this we need the Fisher information and the variance of the estimator. Here the estimator is $\bar{Z}$ and the model is the sample $(X_1, \dots, X_n)$. Thus the Fisher information is calculated w.r.t the law of the vector $(X_1, \dots, X_n)$. The Fisher information is additive so that the Fisher information of $(X_1, \dots, X_n)$ is simply ni_λ where i_λ is the Fisher information of a singe X. The loglikelihood

$$l(\lambda) = -\lambda - \log(1-e^{-\lambda}) + x\log\lambda - \log x!$$

and

$$\begin{aligned}\frac{\partial l}{\partial \lambda} &= -\frac{1}{1-e^{-\lambda}} + \frac{x}{\lambda} \\ \frac{\partial^2 l}{\partial \lambda^2} &= \frac{e^{-\lambda}}{(1-e^{-\lambda})^2} - \frac{x}{\lambda^2}\end{aligned}$$

The Fisher information for one observation is (using $E[X] = \lambda/(1-e^{-\lambda})$)

$$\begin{aligned}i_\lambda &= -\mathbb{E}\left(\frac{\partial^2 l}{\partial\lambda^2}\right) \\ &= -\frac{e^{-\lambda}}{(1-e^{-\lambda})^2} + \frac{1}{\lambda^2}\frac{\lambda}{1-e^{-\lambda}} \\ &= \frac{1}{\lambda}\cdot\frac{1}{1-e^{-\lambda}}\left[1 - \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right]\end{aligned}$$

George Deligiannidis: deligian@stats.ox.ac.uk

To obtain the variance consider

$$\mathbb{E}[Z(Z-1)] = \sum_{x\geq 2} x(x-1)\frac{e^{-\lambda}}{1-e^{-\lambda}}\frac{\lambda^x}{x!} = \frac{e^{-\lambda}}{1-e^{-\lambda}}\sum_{x\geq 2}\frac{\lambda^x}{(x-2)!} = \frac{\lambda^2}{1-e^{-\lambda}}$$

Then

$$\mathrm{Var}(Z) = \frac{\lambda^2}{1-e^{-\lambda}} + \lambda - \lambda^2 = \lambda\left[1 + \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right]$$

I have

$$\begin{aligned} i_\lambda &= -\mathbb{E}\left(\frac{\partial^2 l}{\partial\lambda^2}\right) \\ &= -\frac{e^{-\lambda}}{(1-e^{-\lambda})^2} + \frac{\lambda}{\lambda^2} \\ &= \frac{1}{\lambda}\cdot\frac{1}{1-e^{-\lambda}}\left[1 - \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right] \end{aligned}$$

$$\begin{aligned} \text{Efficiency} &= \left[I_\lambda \mathrm{Var}(\bar{Z})\right]^{-1} \\ &= \frac{1-\left(\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right)^2}{1-e^{-\lambda}}. \end{aligned}$$

Section C

6. (a) (optional bookwork) Let X be a discrete random variable with pmf $f(x;\theta)$ with parameter $\theta \in \Theta$ and sample space $X \in \chi$. Let $T(x)$ be a function of x. Suppose $f(x;\theta)/f(y;\theta)$ is not a function of θ if and only if $T(x) = T(y)$. Show that $T(x)$ is minimal sufficient for θ.

   (b) Let $N = N(0, S]$ be the number of events in a Poisson arrival process of rate λ acting over time s in the interval $0 < s \leq S$. Suppose we observe arrivals in the process at times $X_1, X_2, ..., X_N$, and wish to use these data to estimate λ. Show that N is minimal sufficient for λ (assume the result in (a) holds for any sufficiently regular family of probability distributions).

Solution:

(a) Break the condition into two parts:

(*) $T(x) = T(y) = t$ implies $f(x;\theta)/f(y;\theta)$ is not a function of θ;

(**) $f(x;\theta)/f(y;\theta)$ not a function of θ implies $T(x) = T(y) = t$.

Let $f(x;\theta) = g(x|t(x),\theta)h(t|\theta)$ (with no assumption of sufficiency) and suppose $T(x) = T(y) = t$. If (*) holds then

$$\frac{f(x;\theta)}{f(y;\theta)} = \frac{g(x|t,\theta)}{g(y|t,\theta)} = c(x,y)$$

say, with c independent of θ (factors of h cancel). But then

$$\sum_{x:T(x)=t} g(x|t,\theta) = g(y|t,\theta) \sum_{x:T(x)=t} c(x,y)$$

so

$$g(y|t,\theta) = \left[\sum_{x:T(x)=t} c(x,y)\right]^{-1}$$

which is independent of θ, so T is sufficient for θ in f. If $f(x;\theta)/f(y;\theta)$ does depend on θ when $T(x) = T(y) = t$ then c depends on θ and the same reasoning shows T cannot be sufficient, so condition (*) is necessary for sufficiency. Let $U(x)$ be some sufficient statistic. We must show that T is a function of U, so T is minimal. It is enough to show that $U(x) = U(y)$ implies $T(x) = T(y)$. But $U(x) = U(y) = u$ implies $f(x;\theta)/f(y;\theta)$ is not a function of θ, and then $(**)$ implies $T(x) = T(y)$, so T is minimal sufficient.

George Deligiannidis: deligian@stats.ox.ac.uk

(b) The intervals of a Poisson arrival process of rate λ are exponential so $X_i \sim \mathrm{Exp}(\lambda)$ likelihood for $i = 1, 2, ..., N$. The probability that the final interval between time $Y = \sum_{i=1}^{N} X_i$ and S has no event is the probability that an $\mathrm{Exp}(\lambda)$ random variable exceeds $S - Y$, that is, $\exp(-\lambda(S - Y))$. The likelihood for λ given data $X = (x_1, ...x_n)$ is therefore

$$\begin{aligned} L(\theta; x) &= \left[\prod_{i=1}^{n} \lambda \exp(-\lambda x_i)\right] \exp(-\lambda(S - Y)) \\ &= \exp(-\lambda S)\lambda^n \end{aligned}$$

since $(S - Y) + x_n + ... + x_1 = S$ and so N is sufficient for λ by the factorization theorem ($L = K_1(x, \theta)K_2(x)$ with $K_1(x, \theta) = L$ and $K_2 = 1$). It is minimal sufficient by part (a) since, if $x = (x_1, ...x_n)$ and $y = y_1, ..., y_m$ then $L(x; \lambda)/L(y; \lambda)$ is independent of λ if and only if $n = m$.

7. A random sample $X_1, \ldots, X_n$ is taken from the Weibull distribution

$$\frac{\beta}{\alpha^\beta} x^{\beta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^\beta\right\},\ x > 0, \alpha > 0, \beta > 0.$$

(a) Assuming that β is known, find a sufficient statistic for α.

(b) Suppose now that α is known. Show that the order statistics $X_{(1)}, \ldots, X_{(n)}$ is sufficient statistic for β, but that no one-dimensional statistic can be sufficient.

(c) Does the Weibull distribution belong to a 2-parameter exponential family?

Solution:

$$L(\theta; \mathbf{x}) = \alpha^{-n\beta} \exp\{-\alpha^{-\beta} \sum_{1}^{n} x_i^\beta\} \times \beta^n \prod_{1}^{n} x_i^{\beta-1}.$$

Assuming β is a known constant, this is exponential form in the natural parameter $-\alpha^{-\beta}$. The natural observation $T(x) = n^{-1} \sum_1^n x_i^\beta$ is thus a (minimal) sufficient statistic for α if β is known.

We suppose now that α is known. Observe that the order statistic is always sufficient when the observation is an i.i.d. sample (the order in which the observations arrive contains no information).

Notice that a statistic T is minimal sufficient if and only if $T(x) = T(y)$ is equivalent to $f(x; \theta)/f(y; \theta)$ being independent of θ. In the case of the Weibull distribution, say with α known, and n i.i.d. observations, the log-likelihood ratio takes the form

$$\begin{aligned}F(\boldsymbol{x}, \boldsymbol{y}; \beta) &:= \log \frac{f(x_1, \ldots, x_n; \beta)}{f(y_1, \ldots, y_n; \beta)} \\ &= (\beta - 1) \sum \log(x_i) - \sum \left(\frac{x_i}{\alpha}\right)^\beta - (\beta - 1) \sum \log(y_i) + \sum \left(\frac{y_i}{\alpha}\right)^\beta\end{aligned}$$

and this should be independent of β. For $\beta = 1$ the above implies that $\sum x_i = \sum y_i$.

In addition all the derivatives of the above expression w.r.t. β must vanish. Writing $w_i = \log(x_i/\alpha)$, $z_i = \log(y_i/\beta)$ we have for $p \geq 2$

$$\frac{\partial^p}{\partial \beta^p} F(\beta) = -\sum w_i^p \mathrm{e}^{\beta w_i} + \sum z_i^p \mathrm{e}^{\beta z_i} = 0$$

for all $\beta > 0$. Letting $\beta \to 0$ we obtain then that

$$\sum w_i^p = \sum z_i^p,$$

and therefore all moments of the empirical measures

$$\sum_{i=1}^n \delta_{w_i}, \sum_{i=1}^n \delta_{z_i},$$

are the same and we can conclude that

$$\{x_1, \ldots, x_n\} = \{y_1, \ldots, y_n\}.$$

Therefore $f(\boldsymbol{x}; \beta)/f(\boldsymbol{y}; \beta)$ being independent of θ is equivalent to $\boldsymbol{x}$ being equal to $\boldsymbol{y}$ up to permutation. Therefore the order statistic is minimal sufficient; in particular as n grows so does the dimension of any sufficient statistic. A 2-parameter exponential family admits a 2-dimensional sufficient statistic independent of the size of the sample (see Corollary 2.3 and the remark thereafter), thus giving us a contradiction.

George Deligiannidis: deligian@stats.ox.ac.uk