Fundamental Theory of Statistical Inference
Problems, with solutions

G. A. Young

February 2013
Problems

2.1 Let $X$ be uniformly distributed on $[0, \theta]$ where $\theta \in (0, \infty)$ is an unknown parameter. Let the action space be $[0, \infty)$ and the loss function $L(\theta, d) = (\theta - d)^2$ where $d$ is the action chosen. Consider the decision rules $d_\mu(x) = \mu x, \mu \geq 0$. For what value of $\mu$ is $d_\mu$ unbiased? Show that $\mu = 3/2$ is a necessary condition for $d_\mu$ to be admissible.

[A2.1 When working with the quadratic loss function, a decision rule $d$ is unbiased if and only if $E(d(X)) = \theta$. When considering decision rules $d_\mu$ of the form given, we require $\mu E(X) = \theta$. But, $E(X) = \theta/2$, so $d_\mu$ is unbiased if and only if $\mu = 2$.

Note that

$$R(\theta, d_\mu) = \int_0^\theta (\theta - \mu x)^2 \frac{1}{\theta} dx = (1 - \mu + \mu^2/3)\theta^2.$$  

The quadratic function of $\mu$ has a unique minimum at $\mu = 3/2$: if $\mu \neq 3/2$, then $d_\mu$ is strictly dominated by $d_{3/2}$, so $\mu = 3/2$ is a necessary condition for admissibility.]

2.3 Each winter evening between Sunday and Thursday, the superintendent of the Chapel Hill School District has to decide whether to call off the next day’s school because of snow conditions. If he fails to call off school and there is snow, there are various possible consequences, including children and teachers failing to show up for school, the possibility of traffic accidents etc. If he calls off school, then regardless of whether there actually is snow that day, there will have to be a make-up day later in the year. After weighing up all the possible outcomes he decides that the costs of failing to close school when there is snow are twice the costs incurred by closing school, so he assigns two units of loss to the first outcome and one to the second. If he does not call off school and there is no snow, then of course there is no loss.

Two local radio stations give independent and identically distributed weather forecasts. If there is to be snow, each station will forecast this with probability $3/4$, but predict no snow with probability $1/4$. If there is to be no snow, each station predicts snow with probability $1/2$.

The superintendent will listen to the two forecasts this evening, and then make his decision on the basis of the data $x$, the number of stations forecasting snow.

Write down an exhaustive set of non-randomised decision rules based on $x$. 

Find the superintendent’s admissible decision rules, and his minimax rule. Before listening to the forecasts, he believes there will be snow with probability 1/2; find the Bayes rule with respect to this prior.

[Include randomised rules in your analysis when determining admissible, minimax and Bayes rules].

**A2.3** There are two actions: C, ‘close school’, and K, ‘keep open school’. Two states of nature, $\theta = 0$, ‘no blizzard’, and $\theta = 1$, ‘blizzard’. Let $x$ be number of radio stations predicting blizzard. There are 8 non-randomised decision rules, as described below:

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_5$</th>
<th>$d_6$</th>
<th>$d_7$</th>
<th>$d_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 0$</td>
<td>C</td>
<td>K</td>
<td>C</td>
<td>K</td>
<td>C</td>
<td>K</td>
<td>C</td>
</tr>
<tr>
<td>$x = 1$</td>
<td>C</td>
<td>C</td>
<td>K</td>
<td>K</td>
<td>C</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>$x = 2$</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>K</td>
<td>K</td>
<td>K</td>
</tr>
</tbody>
</table>

$R(0, \cdot)$

$\begin{array}{cccccc}
1 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\
\end{array}$

$R(1, \cdot)$

$\begin{array}{cccccc}
1 & \frac{17}{16} & \frac{22}{16} & \frac{23}{16} & \frac{25}{16} & \frac{26}{16} & \frac{31}{16} & 2 \\
\end{array}$

Draw the associated risk set.
The admissible rules are those on the ‘south west boundary’ of the risk set: any convex combination of $d_3$ and $d_4$, or $d_4$ and $d_2$, or $d_2$ and $d_1$. The minimax rule (corresponding to the intersection of $[0,k]^2$ with the risk set for minimal $k$, here $k = 1$) is $d_4$. The Bayes rule with respect to the given prior (corresponding to the point of intersection of $\frac{1}{2}R(0,d) + \frac{1}{2}R(1,d) = k$ with the risk set for minimal $k$) is $d_4$.]

2.5 Bacteria are distributed at random in a fluid, with mean density $\theta$ per unit volume, for some $\theta \in H \subseteq [0, \infty)$. This means that

$$\Pr_\theta(\text{no bacteria in volume } v) = e^{-\theta v}.$$ 

We remove a sample of volume $v$ from the fluid and test it for the presence or absence of bacteria. On the basis of this information we have to decide whether there are any bacteria in the fluid at all. An incorrect decision will result in a loss of 1, a correct decision in no loss.

(i) Suppose $H = [0, \infty)$. Describe all the non-randomised decision rules for this problem and calculate their risk functions. Which of these rules are admissible?

(ii) Suppose $H = \{0, 1\}$. Identify the risk set

$$S = \{(R(0,d), R(1,d)) : d \text{ a randomised rule} \} \subseteq \mathbb{R}^2,$$

where $R(\theta,d)$ is the expected loss in applying $d$ under $\Pr_\theta$. Determine the minimax rule.

(iii) Suppose again that $H = [0, \infty)$.

Determine the Bayes decision rules and Bayes risk for prior

$$\pi(\{0\}) = \frac{1}{3},$$
$$\pi(A) = \frac{2}{3} \int_A e^{-\theta} d\theta, \quad A \subseteq (0, \infty).$$

[So the prior probability that $\theta = 0$ is 1/3, while the prior probability that $\theta \in A \subseteq (0, \infty)$ is $2/3 \int_A e^{-\theta} d\theta$.]

(iv) If it costs $v/24$ to test a sample of volume $v$, what is the optimal volume to test? What if the cost is 1/6 per unit volume?

[A2.5 (i) There are two actions: N ‘decide the fluid is not contaminated with bacteria’, C ‘decide fluid is contaminated’. Test result is $x$, say, with $x = 0$ ‘no bacteria in sample’, $x = 1$ ‘bacteria in sample’.

There are four non-randomised decision rules:
<table>
<thead>
<tr>
<th></th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 0$</td>
<td>$N$</td>
<td>$N$</td>
<td>$C$</td>
<td>$C$</td>
</tr>
<tr>
<td>$x = 1$</td>
<td>$N$</td>
<td>$C$</td>
<td>$N$</td>
<td>$C$</td>
</tr>
</tbody>
</table>

$R(\theta = 0, \cdot)$

- $0$
- $0$
- $1$
- $1$

$R(\theta > 0, \cdot)$

- $e^{-\theta v}$
- $1 - e^{-\theta v}$
- $0$

Plot the risk functions on $[0, \infty)$ to see that $d_1$ is strictly dominated by $d_2$ and $d_3$ is strictly dominated by $d_4$: both $d_2$ and $d_4$ are admissible.

(ii) Draw the risk set.

The minimax rule is a randomised rule, of the form $d = pd_2 + (1 - p)d_4$, and is identified by setting $R(\theta = 0, d) = R(\theta = 1, d)$. Then $p = 1/(1 + e^{-v})$ is easily calculated.

(iii) Only need to consider $d_2$ and $d_4$. We have

$$r(\pi, d_2) = \frac{2}{3} \int_0^\infty e^{-\theta v} e^{-\theta v} d\theta = \frac{2}{3(1 + v)}$$

$$r(\pi, d_4) = \frac{1}{3}$$

Then $d_2$ is Bayes if $v > 1$, $d_4$ is Bayes if $v < 1$, with any randomised rule formed by the two Bayes if $v = 1$. 
(iv) Let the Bayes risk of testing a sample of volume \( v \) be \( r(v) \), and let cost of sampling be \( c \) per unit volume. Total cost is \( cv + r(v) \). This is \( cv + 1/3 \) if \( v \leq 1 \) and is \( cv + 2/\{3(1 + v)\} \) if \( v \geq 1 \). Careful minimisation with respect to \( v \) (remembering that the minimum might occur at \( v = 0 \)) gives optimal \( v = 3 \) if \( c = 1/24 \) and \( v = 0 \) if \( c = 1/6 \).]

2.6 Prove Theorems 2.3, 2.4 and 2.5, concerning admissibility of Bayes rules.

[A2.6 For 2.3: suppose that \( d \) is Bayes with respect to \( \pi \), and suppose for a contradiction that \( d' \) strictly dominates \( d \). Then, \( R(\theta, d') \leq R(\theta, d) \) for all \( \theta \in \Omega_\theta \), and \( R(\theta_j, d') < R(\theta_j, d) \) for some \( j = 1, \ldots, t \). But then, since the prior attaches positive probability to each \( \theta_j \), we would have

\[
r(\pi, d') < r(\pi, d),
\]

contradicting the fact that \( d \) is a Bayes rule with respect to \( \pi \).

For 2.4: suppose that \( d \) is the unique Bayes rule with respect to the prior \( \pi \), and that \( d' \) strictly dominates \( d \), so that \( d \) is inadmissible. Then

\[
r(\pi, d') = \int_{\theta \in \Omega_\theta} R(\theta, d') \pi(\theta) d\theta \leq \int_{\theta \in \Omega_\theta} R(\theta, d) \pi(\theta) d\theta = r(\pi, d),
\]

contradicting the fact that \( d \) is the unique Bayes rule with respect to \( \pi \), so \( d \) is admissible.

For 2.5: suppose that \( d \) is a Bayes rule with respect to \( \pi \), but is inadmissible. Then there is a \( d' \) that strictly dominates \( d \). Let \( \theta_0 \) be a value of \( \theta \) at which the risk of \( d' \) is strictly less than that of \( d \), and suppose \( R(\theta_0, d) - R(\theta_0, d') = \epsilon \), for \( \epsilon > 0 \). Since the risk function is continuous in \( \theta \) for any decision rule, there exists \( \delta > 0 \) such that either \( (\theta_0 - \delta, \theta_0) \) or \( (\theta_0, \theta_0 + \delta) \) is contained in \( \Omega_\theta \), and

\[
R(\theta, d) - R(\theta, d') > \frac{\epsilon}{2} \text{ for all } |\theta - \theta_0| < \delta.
\]

Then

\[
r(\pi, d) - r(\pi, d') \geq \int_{\theta_0 - \delta}^{\theta_0 + \delta} (R(\theta, d) - R(\theta, d')) \pi(\theta) d\theta > \int_{\theta_0 - \delta}^{\theta_0 + \delta} \frac{\epsilon}{2} \pi(\theta) d\theta > 0,
\]

since the interval \( (\theta_0 - \delta, \theta_0 + \delta) \) has positive probability under \( \pi \). This contradicts \( d \) being Bayes, so it is admissible.]

2.8 In a Bayes decision problem, a prior distribution \( \pi \) is said to be least favourable if \( r_\pi \geq r_{\pi'} \), for all prior distributions \( \pi' \), where \( r_\pi \) denotes the Bayes risk of the Bayes rule \( d_\pi \) with respect to \( \pi \).
Suppose that \( \pi \) is a prior distribution, such that
\[
\int R(\theta, d_\pi) \pi(\theta) d\theta = \sup_\theta R(\theta, d_\pi).
\]

Show that (i) \( d_\pi \) is minimax, (ii) \( \pi \) is least favourable.

[A2.8] Let \( \delta \) be any other decision rule, then
\[
\sup_\theta R(\theta, \delta) \geq \int R(\theta, \delta) \pi(\theta) d\theta \geq \int R(\theta, d_\pi) \pi(\theta) d\theta,
\]
be the definition of \( d_\pi \). The latter quantity is \( \sup_\theta R(\theta, d_\pi) \) by assumption, so (i) is proved. Let \( \pi' \) be any other prior. Then
\[
r_{\pi'} = \int R(\theta, d_{\pi'}) \pi'(\theta) d\theta \leq \int R(\theta, d_{\pi}) \pi'(\theta) d\theta,
\]
since \( d_{\pi'} \) is Bayes with respect to \( \pi' \). But the latter quantity \( \leq \sup_\theta R(\theta, d_{\pi'}) \equiv r_{\pi} \), by assumption, proving (ii).]

3.3 Find the form of the Bayes rule in an estimation problem with loss function
\[
L(\theta, d) = \begin{cases} 
   a(\theta - d) & \text{if } d \leq \theta \\
   b(d - \theta) & \text{if } d > \theta,
\end{cases}
\]
where \( a \) and \( b \) are given positive constants.

[A3.3] We have that \( d \) minimizes the expected posterior loss
\[
a \int_{-\infty}^{d} (\theta - d) \pi(\theta \mid x) d\theta + b \int_{d}^{\infty} (d - \theta) \pi(\theta \mid x) d\theta.
\]

Differentiate with respect to \( d \) to see that \( d \) is \( a/(a + b) \) quantile of posterior \( \pi(\theta \mid x) \):
\[
\int_{-\infty}^{d} \pi(\theta \mid x) d\theta = \frac{a}{a + b}.
\]

3.4 Suppose that \( X \) is distributed as a binomial random variable with index \( n \) and parameter \( \theta \). Calculate the Bayes rule (based on the single observation \( X \)) for estimating \( \theta \) when the prior distribution is the uniform distribution on \([0, 1]\) and the loss function is
\[
L(\theta, d) = (\theta - d)^2 / \{\theta(1 - \theta)\}.
\]
Is the rule you obtain minimax?

[A3.4 The posterior density of $\theta$ given $X = x$ has

$$\pi(\theta|x) \propto \theta^x(1-\theta)^{n-x}.$$  

The Bayes rule minimizes the expected posterior loss

$$\int_0^1 (\theta - d)^2 \theta^{x-1}(1-\theta)^{n-x-1}d\theta.$$  

Differentiation with respect to $d$ shows that the minimum occurs when $d \equiv d(x) = x/n$. So the Bayes rule is $d(X) = X/n$. Direct calculation of the risk,

$$R(\theta, d) = E\{(\theta - X/n)^2\}/\{\theta(1-\theta)\} = var(X)/\{(n^2\theta(1-\theta)) = 1/n,$$

shows that the risk is constant, so the rule is minimax.]

3.5 At a critical stage in the development of a new aeroplane, a decision must be taken to continue or to abandon the project. The financial viability of the project can be measured by a parameter $\theta$, $0 < \theta < 1$, the project being profitable if $\theta > \frac{1}{2}$. Data $x$ provide information about $\theta$.

If $\theta < \frac{1}{2}$, the cost to the taxpayer of continuing the project is $(\frac{1}{2} - \theta)$ [in units of $\$\text{billion}$], whereas if $\theta > \frac{1}{2}$ it is zero (since the project will be privatised if profitable). If $\theta > \frac{1}{2}$ the cost of abandoning the project is $(\theta - \frac{1}{2})$ (due to contractural arrangements for purchasing the aeroplane from the French), whereas if $\theta < \frac{1}{2}$ it is zero. Derive the Bayes decision rule in terms of the posterior mean of $\theta$ given $x$.

The Minister of Aviation has prior density $6\theta(1-\theta)$ for $\theta$. The Prime Minister has prior density $4\theta^3$. The prototype aeroplane is subjected to trials, each independently having probability $\theta$ of success, and the data $x$ consist of the total number of trials required for the first successful result to be obtained. For what values of $x$ will there be serious ministerial disagreement?

[A3.5 For the given loss function, the Bayes rule is to abandon if

$$\int_0^\frac{1}{2} (\frac{1}{2} - \theta)\pi(\theta|x)d\theta > \int_{\frac{1}{2}}^1 (\theta - \frac{1}{2})\pi(\theta|x)d\theta,$$

which reduces to $\mu(x) \equiv \int_0^x \theta\pi(\theta|x)d\theta < \frac{1}{2}$.  

MA and PM both have beta priors, and $f(x; \theta) = \theta(1-\theta)^{x-1}, x = 1, 2, ....$, so their two posteriors are both of beta form also. The Beta$(\alpha, \beta)$ distribution
has mean $a/(a + b)$, so the MA will wish to abandon if $3/(4 + x) < 1/2$, or $x > 2$, with 2 borderline. PM abandons if $5/(5 + x) < 1/2$, or $x > 5$, with 5 borderline. Serious disagreement if $x = 3$ or 4.

3.8 Suppose $X_1, \ldots, X_n$ are independent, identically distributed random variables which, given $\mu$, have the normal distribution $N(\mu, \sigma^2_0)$, with $\sigma^2_0$ known. Suppose also that the prior distribution of $\mu$ is normal with known mean $\xi_0$ and known variance $\nu_0$.

Let $X_{n+1}$ be a single future observation from the same distribution which is, given $\mu$, independent of $X_1, \ldots, X_n$. Show that, given $(X_1, \ldots, X_n)$, $X_{n+1}$ is normally distributed with mean

$$\left\{ \frac{1}{\sigma^2_0/n + \nu_0} \right\}^{-1} \left\{ \frac{\overline{X}}{\sigma^2_0/n + \nu_0} + \frac{\xi_0}{\nu_0} \right\}$$

and variance

$$\sigma^2_0 + \left\{ \frac{1}{\sigma^2_0/n + \nu_0} \right\}^{-1}.$$

[A3.8 This is a matter of algebra, with the usual trick of completing the square. Both the posterior density $p(\mu | X_1, \ldots, X_n)$ and the density of $X_{n+1}$ given $\mu$ are normal: to obtain the predictive distribution we must integrate the product.]

3.9 Let $X_1, \ldots, X_n$ be independent, identically distributed $N(\mu, \sigma^2)$, with both $\mu$ and $\sigma^2$ unknown. Let $\overline{X} = n^{-1} \sum_{i=1}^n X_i$, and $s^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \overline{X})^2$.

Assume the (improper) prior $p(\mu, \sigma)$ with

$$p(\mu, \sigma) \propto \sigma^{-1}, (\mu, \sigma) \in \mathbb{R} \times (0, \infty).$$

Show that the marginal posterior distribution of $n^{1/2}(\mu - \overline{X})/s$ is the t distribution with $n - 1$ degrees of freedom, and find the marginal posterior distribution of $\sigma$.

[A3.9 We have that the likelihood function is

$$f(x; \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{\nu s^2}{2\sigma^2} - \frac{n(\overline{X} - \mu)^2}{2\sigma^2} \right\},$$

with $\nu = n - 1$.

The joint posterior is the product of the likelihood and the prior:

$$p(\mu, \sigma | x) \propto \frac{1}{\sigma^{n+1}} \exp \left[ -\frac{\nu s^2(1 + n(\overline{X} - \mu)^2)}{2\sigma^2} \right].$$]
Integrate over $\sigma$ to obtain

$$\pi(\mu | x) \propto \left\{1 + \frac{n (\bar{X} - \mu)^2}{\nu s^2}\right\}^{-n/2},$$

which gives the required conclusion. Integrate out $\mu$ from the joint posterior to obtain

$$\pi(\sigma | x) \propto \frac{1}{\sigma^{\nu+1}} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right), \quad \sigma > 0.$$  

This is expressed as $\sigma|x \sim \nu^{1/2} s \chi_\nu^{-1}$, where $\chi_\nu^{-1}$ is a random variable with the 'inverse $\chi$ distribution with $\nu$ degrees of freedom'.

The analytic key here is to note that

$$\int_0^\infty x^\kappa \exp(-\lambda x) dx = \frac{1}{\alpha \lambda \kappa + 1} \Gamma(\kappa/\alpha)\!.$$  

\[3.10\] Consider a Bayes decision problem with scalar parameter $\theta$. An estimate is required for $\phi \equiv \phi(\theta)$, with loss function

$$L(\theta, d) = (d - \phi)^2.$$  

Find the form of the Bayes estimator of $\phi$.

Let $X_1, \ldots, X_n$ be independent, identically distributed random variables from the density $\theta e^{-\theta x}, x > 0$, where $\theta$ is an unknown parameter. Let $Z$ denote some hypothetical future value derived from the same distribution, and suppose we wish to estimate $\phi(\theta) = \Pr(Z > z)$, for given $z$.

Suppose we assume a gamma prior, $\pi(\theta) \propto \theta^{\alpha-1} e^{-\theta \beta}$ for $\theta$. Find the posterior distribution for $\theta$, and show that the Bayes estimator of $\phi$ is

$$\tilde{\phi}_B = \left(\frac{\beta + S_n}{\beta + S_n + z}\right)^{\alpha + n},$$

where $S_n = X_1 + \ldots + X_n$.

\[A3.10\] The standard argument shows that $d(x)$ minimises the expected posterior loss

$$\int_{\Theta} L(\theta, d) \pi(\theta | x) d\theta = \int_{\Theta} (d - \phi(\theta))^2 \pi(\theta | x) d\theta.$$  

Differentiation shows $d = \int_{\Theta} \phi(\theta) \pi(\theta | x) d\theta$, the posterior mean of $\phi(\theta)$. 
In the example, $\phi = e^{-\theta^2}$, with the posterior being $\text{Gamma}(\alpha + n, \beta + S_n)$. To verify the Bayes estimator, note that this is
\[
\int_0^\infty e^{-\theta^2} \pi(\theta|x) d\theta = \frac{(\beta + S_n)^{\alpha+n}}{(\beta + S_n + z)^{\alpha+n}} \int_0^\infty (\beta + S_n + z)^{\alpha+n} \theta^{\alpha+n-1} e^{-(\theta + S_n + z)\theta} \Gamma(\alpha + n) d\theta
\]
\[
= \left( \frac{\beta + S_n}{\beta + S_n + z} \right)^{\alpha+n},
\]
on noting that the integrand is a gamma pdf.]

3.11 Let the distribution of $X$, given $\theta$, be normal with mean $\theta$ and variance $1$. Consider estimation of $\theta$ with squared error loss $L(\theta, \alpha) = (\theta - \alpha)^2$ and action space $\Lambda \equiv \Omega_\theta \equiv \mathbb{R}$.

Show that the usual estimate of $\theta$, $d(X) = X$, is not a Bayes rule.

[Show that if $d(X)$ were Bayes with respect to a prior distribution $\pi$, we should have $r(\pi, d) = 0$.]

Show that $X$ is extended Bayes and minimax.

A3.11 Squared error loss, so if $d$ were Bayes, we should have $d(X) = E(\theta|X)$. But then,
\[
E\{\theta d(X)\} = E[E\{\theta d(X)|X\}] = E[d(X) E(\theta|X)] = E[d(X)^2].
\]
Also, $E[\theta d(X)] = E[E\{\theta d(X)|\theta\}] = E(\theta^2)$, giving $r(\pi, d) = E(\theta^2) - 2E[\theta d(X)] + E[d(X)^2] = 0$. But, $r(\pi, d)$ is actually
\[
r(\pi, d) = E[\theta - d(X)]^2 = E[E\{(\theta - X)^2|\theta\}] = E[\text{var}(X|\theta)] = 1,
\]
a contradiction.

Consider a $N(0, \sigma^2)$ prior, $\pi_\sigma$ say. The posterior is $N(x\sigma^2/(1 + \sigma^2), \sigma^2/(1 + \sigma^2))$. The Bayes rule with respect to $\pi_\sigma$ is $d_\sigma(x) = x\sigma^2/(1 + \sigma^2)$, with Bayes risk $r(\pi_\sigma, d_\sigma) = \sigma^2/(1 + \sigma^2) = \inf_\delta r(\pi_\sigma, \delta)$. Then
\[
r(\pi_\sigma, d) = 1 = \inf_\delta r(\pi_\sigma, \delta) + \epsilon,
\]
for $\epsilon = 1/(1 + \sigma^2)$. By suitable choice of $\sigma$ we have that $d$ is $\epsilon$-Bayes for every $\epsilon > 0$, i.e. extended Bayes. Since $d$ is also an equalizer rule, it is minimax.

5.1 Prove that random samples from the following distributions form $(m, m)$ exponential families with either $m = 1$ or $m = 2$: Poisson, binomial, geometric, gamma (index known), gamma (index unknown). Identify
the natural statistics and the natural parameters in each case. What are the distributions of the natural statistics?

The negative binomial distribution with both parameters unknown provides an example of a model that is not of exponential family form. Why?

[If Y has a gamma distribution of known index k, its density function is of the form

\[ f_Y(y; \mu) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{\Gamma(k)}. \]

The gamma distribution with index unknown has both k and \( \lambda \) unknown.]

[A5.1 Just a matter of rewriting the density functions in exponential family form. With the negative binomial distribution, the sample space depends on the unknown parameter, not allowed in exponential family, if both parameters are unknown.]

5.2 Let \( Y_1, \ldots, Y_n \) be independent, identically distributed \( N(\mu, \mu^2) \).

Show that this model is an example of a curved exponential family.

[A5.2 Immediate from writing joint density in exponential family form:

\[ P_Y(y; \mu) \propto \exp \left\{ -\frac{1}{2\mu^2} \sum y_i^2 + \frac{1}{\mu} \sum y_i - n \log \mu \right\}. \]

]

5.3 Find the general form of a conjugate prior density for \( \theta \) in a Bayesian analysis of the one-parameter exponential family density

\[ f(x; \theta) = c(\theta) h(x) \exp(\theta t(x)), \ x \in \mathbb{R}. \]

[A5.3 Suppose that the prior density is of the form

\[ \pi(\theta) \propto c(\theta)^a \exp(t_0 \theta), \]

for constants \( a, t_0 \). Then the posterior density \( \pi(\theta|x) \) has

\[ \pi(\theta|x) \propto c(\theta)^{a+1} \exp\{(t_0 + t(x))\theta\}. \]

]

5.4 Verify that the family of gamma distributions of known index constitutes a transformation model under the action of the group of scale transformations.
[This provides an example of a family of distributions which constitutes both an exponential family, and a transformation family. Are there any others?]

[A5.4] Let $Y \sim \text{Gamma}(k, \lambda)$ denote $Y$ has a gamma distribution of known index $k$ and pdf

$$f_Y(y; \lambda) = \frac{\lambda^k y^{k-1}e^{-\lambda y}}{\Gamma(k)}.$$ 

Then $\sigma Y \sim \text{Gamma}(k, \lambda/\sigma)$.

5.5 The maximum likelihood estimator $\hat{\theta}(x)$ of a parameter $\theta$ maximises the likelihood function $L(\theta) = f(x; \theta)$ with respect to $\theta$. Verify that maximum likelihood estimators are equivariant with respect to the group of one-to-one transformations.

[A5.5] This follows from the transformation property, that if $\phi = \phi(\theta)$ is a one-to-one transformation of the parameter $\theta$, then $\phi = \phi(\bar{\theta})$, where $\bar{\theta} = s(Y)$ is the maximum likelihood estimator of $\theta$. If the transformation $\phi(\cdot)$ corresponds to $g_\phi \in G$, then $g_\phi (Y)$ is the transformation of $Y$ whose MLE is $\phi$. Then $\tilde{\phi} = s(g_\phi (Y))$, while $\phi(\tilde{\theta}) = \bar{g}_\phi (s(Y))$. Hence $s(g_\phi (Y)) = \bar{g}_\phi (s(Y))$, for all such $g_\phi$, which is the requirement of equivariance.

5.6 Verify directly that in the location-scale model the configuration has a distribution which does not depend on the parameters.

[A5.6] This is just a matter of writing the configuration in terms of $\varepsilon_1, \ldots, \varepsilon_n$, where $Y_j = \eta + \tau \varepsilon_j$.

6.1 Let $X_1, \ldots, X_n$ be independent, identically distributed $N(\mu, \sigma^2)$ random variables.

Find a minimal sufficient statistic for $\mu$ and show that it is not complete.

[A6.1] By examination of the likelihood ratio, we see $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is minimal sufficient. Show directly that

$$E \left( \frac{(\sum_{i=1}^n X_i)^2}{1 + n} - \sum_{i=1}^n X_i^2 \right) = 0,$$

to show that it is not complete.

6.2 Find a minimal sufficient statistic for $\theta$ based on an independent sample of size $n$ from each of the following distributions:

(i) the gamma distribution with density

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0,$$
with \( \theta = (\alpha, \beta) \);

(ii) the uniform distribution on \( (\theta - 1, \theta + 1) \);

(iii) the Cauchy distribution with density

\[
f(x; \alpha, b) = \frac{b}{\pi\{(x - \alpha)^2 + b^2\}}, \ x \in \mathbb{R},
\]

with \( \theta = (a, b) \).

[A6.2 For (i) the minimal sufficient statistic is \( \sum X_i, \sum \log(X_i) \). Part (ii) is essentially done in the notes: \( \min X_i, \max X_i \) is minimal sufficient. Part (iii) requires more care. Note that the condition

\[
\frac{L(\theta_1; x)}{L(\theta_2; x)} = \frac{L(\theta_1; y)}{L(\theta_2; y)}, \ \forall \theta_1, \theta_2 \in \Omega_\theta,
\]

means that \( \frac{L(\theta; x)}{L(\theta; y)} \) does not depend on \( \theta \). In the Cauchy case this ratio is a ‘rational function’ of \( \theta \). For the ratio to be independent of \( \theta \) all powers of \( \theta \) must have identical coefficients in the numerator and denominator. This happens if and only if \( (y_1, \ldots, y_n) \) is a permutation of \( (x_1, \ldots, x_n) \), so that the minimal sufficient statistic is the set of order statistics \( (X_{(1)}, \ldots, X_{(n)}) \).

[6.3 Independent factory-produced items are packed in boxes each containing \( k \) items. The probability that an item is in working order is \( \theta \), \( 0 < \theta < 1 \). A sample of \( n \) boxes are chosen for testing, and \( X_i \), the number of working items in the \( i \)th box, is noted. Thus \( X_1, \ldots, X_n \) are a sample from a binomial distribution, \( \text{Bin}(k, \theta) \), with index \( k \) and parameter \( \theta \). It is required to estimate the probability, \( \theta^k \), that all items in a box are in working order. Find the minimum-variance unbiased estimator, justifying your answer.

[A6.3 We have \( S_n = X_1 + \ldots + X_n \) is a sufficient statistic and \( S_n/nk \) is an unbiased estimator of \( \theta \), but \( (S_n/nk)^k \) is not unbiased for \( \theta^k \). We need to find some unbiased estimator: one such is \( I(X_1 = k) \) (\( I := \) indicator function). So the MVUE when \( S_n = t \) will be \( P(X_1 = k|S_n = t) \). Since \( S_n \) is also a complete sufficient statistic, it will follow that this is the unique MVUE. In particular, if we started with a different unbiased estimator of \( \theta^k \), we would end up with the same MVUE. So it does not matter that the actual unbiased estimator we have used is extremely crude.

Now, we can easily check that \( P(X_1 = k|S_n = t) = \binom{n-k}{t-k}/\binom{n}{t} \) (hypergeometric distribution). So the unique MVUE is \( \binom{n-k}{t-k}/\binom{n}{t} \). However, this can lead to a paradoxical estimate: for instance, when \( 0 < S_n < k \) the MVUE is 0 but (because \( S_n > 0 \)) it is impossible that \( \theta^k \) really is 0. ]
6.4 A married man who frequently talks on his mobile is well known to have conversations whose lengths are independent, identically distributed random variables, distributed as exponential with mean \(1/\lambda\). His wife has long been irritated by his behaviour and knows, from infinitely many observations, the exact value of \(\lambda\). In an argument with her husband, the woman produces \(t_1, \ldots, t_n\), the times of \(n\) telephone conversations, to prove how excessive her husband is. He suspects that she has randomly chosen the observations, conditional on their all being longer than the expected length of conversation. Assuming he is right in his suspicion, the husband wants to use the data he has been given to infer the value of \(\lambda\). What is the minimal sufficient statistic he should use? Is it complete? Find the maximum likelihood estimator for \(\lambda\).

[A6.4 Stripped of all the verbiage, this is about the following problem: estimate \(\lambda\) from an independent sample \(T_1, \ldots, T_n\), where the common density is \(f(t; \lambda) = \lambda e^{-\lambda t}, t > 1/\lambda\). The minimal sufficient statistic is \((S_n, Z_n)\) where \(S_n = T_1 + \cdots + T_n\) and \(Z_n = \min\{T_1, \ldots, T_n\}\). However \(E\{S_n\} = 2n\lambda^{-1}\), \(E\{Z_n\} = (1 + 1/n) \lambda^{-1}\), so, for example, \((1 + 1/n) S_n - 2nZ_n\) has mean 0 but is certainly not identically 0. So the minimal sufficient statistic is not complete. The maximum likelihood estimator is 1/\(Z_n\).]

4.1 A random variable \(X\) has one of two possible densities:

\[
 f(x; \theta) = \theta e^{-\theta x}, \quad x \in (0, \infty), \quad \theta \in \{1, 2\}.
\]

Consider the family of decision rules

\[
 d_\mu(x) = \begin{cases} 
 1 & \text{if } x \geq \mu, \\
 2 & \text{if } x < \mu,
\end{cases}
\]

where \(\mu \in [0, \infty]\). Calculate the risk function \(R(\theta, d_\mu)\) for loss function \(L(\theta, d) = |\theta - d|\), and sketch the parametrised curve \(C = \{(R(1, d_\mu), R(2, d_\mu)) : \mu \in [0, \infty]\}\) in \(\mathbb{R}^2\).

Use the Neyman–Pearson Theorem to show that \(C\) corresponds precisely to the set of admissible decision rules.

For what prior mass function for \(\theta\) does the minimax rule coincide with the Bayes rule?

[A4.1 We have \(R(1, d_\mu) = P_1(X < \mu) = 1 - \exp(-\mu)\) and \(R(2, d_\mu) = P_2(X \geq \mu) = \exp(-2\mu) = (1 - R(1, d_\mu))^2\).]
The curve $C$ looks like:

To apply Neyman-Pearson, note that if $\theta = 1$ is null hypothesis and $\theta = 2$ is alternative, then test of a given size (i.e. $R(1, d)$) of maximum power (i.e. minimum $R(2, d)$) is a likelihood ratio test. In this case a likelihood ratio test of a given size is a $d_{\mu}$ for some $\mu$. Formally, NP tells us

$$R(1, d) \leq R(1, \delta_{\mu}) \Rightarrow R(2, d) \geq R(2, \delta_{\mu}).$$

So, the risk set lies above $C$, which therefore represents the admissible decision rules.

$d_{\mu}$ is minimax iff $1 - \lambda = \lambda^2, \lambda = \exp(-\mu)$. The gradient of $C$ at $d_{\mu}$ is $-2\lambda$. Let the prior have $\pi_1 = P(\theta = 1)$ etc. Then $d_{\mu}$ is Bayes iff $2\lambda = \pi_1/\pi_2$. So the minimax and Bayes rules coincide iff $\lambda^2 + \lambda - 1 = 0$ and $2\lambda = \pi_1/\pi_2$: this gives:

$$\pi_1 = 1 - \frac{1}{\sqrt{5}}, \pi_2 = \frac{1}{\sqrt{5}}.$$

4.3 Let $X_1, \ldots, X_n$ be independent random variables with a common density function

$$f(x; \theta) = \theta e^{-\theta x}, x \geq 0,$$

where $\theta \in (0, \infty)$ is an unknown parameter. Consider testing the null hypothesis $H_0 : \theta \leq 1$ against the alternative $H_1 : \theta > 1$. Show how to obtain a uniformly most powerful test of size $\alpha$.

[A4.3 The family is MLR with respect to $t(X) = -\sum X_i$ so the test which rejects $H_0$ when $\sum X_i \leq c$ has increasing power function and is UMP of its size from the general theory. For direct verification of the increasing power function, note that we may define $Z_i = \theta X_i$ so that $\{Z_i\}$ are IID exponential]}
with mean 1. We then have that 
P_\theta(\sum X_i \leq c) = P(\sum Z_i \leq c\theta) = G(c\theta)
where G is the distribution function of \(\sum Z_i\). It follows at once that G(c\theta) is an increasing function of \(\theta\). In fact we know that G is the distribution function of a gamma(\(n, 1\)) random variable so it is possible to write down G directly as an integral of the gamma density.

4.5 Let \(X_1, \ldots, X_n\) be an independent sample of size \(n\) from the uniform distribution on \((0, \theta)\).

Show that there exists a uniformly most powerful size \(\alpha\) test of \(H_0: \theta = \theta_0\) against \(H_1: \theta > \theta_0\), and find its form.

Let \(T = \max(X_1, \ldots, X_n)\).

Show that the test

\[
\phi(x) = \begin{cases} 
  1, & \text{if } t > \theta_0 \text{ or } t \leq b \\
  0, & \text{if } b < t \leq \theta_0,
\end{cases}
\]

where \(b = \theta_0 \alpha^{1/n}\), is a uniformly most powerful test of size \(\alpha\) for testing \(H_0\) against \(H'_1: \theta \neq \theta_0\).

[Note that in a ‘more regular’ situation, a UMP test of \(H_0\) against \(H'_1\) doesn’t exist.]

[A4.5] Consider testing \(H_0: \theta = \theta_0\) against \(H_1: \theta = \theta_1\), where \(\theta_1 > \theta_0\). Neyman Pearson gives most powerful test of form, reject \(H_0\) if \(T > C\), \(T = \max(X_1, \ldots, X_n)\). Under \(H_0\), \(P(T > C) = 1 - \left(\frac{C}{\theta_0}\right)^n\), so \(C = \theta_0(1-\alpha)^{1/n}\) for size \(\alpha\). Test does not depend on \(\theta_1\), so UMP for testing \(H_0: \theta = \theta_0\) against \(H_1: \theta > \theta_0\). Similarly, the UMP test of \(H_0: \theta = \theta_0\) against \(H_1: \theta < \theta_0\) is of the form, reject \(H_0\) if \(T < \theta_0 \alpha^{1/n}\).

So, the maximal power of a size \(\alpha\) test of \(H_0: \theta = \theta_0\) against \(H_1: \theta = \theta_1\) is

\[
1 - \left(\frac{\theta_0}{\theta_1}\right)^n (1 - \alpha), \quad \theta_1 > \theta_0 \\
\left(\frac{\theta_0}{\theta_1}\right)^n \alpha, \quad \theta_1 < \theta_0
\]

The question is completed by showing that for all \(\theta_1\) the given test has power given by \(*\), and that it has size \(\alpha\).

7.1 Let \(X_1, \ldots, X_n\) be an independent sample from a normal distribution with mean 0 and variance \(\sigma^2\). Explain in as much detail as you can how to construct a UMU test of \(H_0: \sigma = \sigma_0\) against \(H_1: \sigma \neq \sigma_0\).

[A7.1] The sufficient statistic is \(\sum X_i^2\). We have \(\sum X_i^2 / \sigma^2 \sim \chi^2_n\). Since the distributions for varying \(\sigma^2\) form an exponential family, a general result
shows that a UMPU test exists, is of two-sided form and has acceptance region of the form

\[ C_1 \leq \Sigma X_i^2 / \sigma_0^2 \leq C_2. \]

The power function \( w(\sigma^2) \) is given by

\[ 1 - w(\sigma^2) = P \left( \frac{C_1 \sigma_0^2 / \sigma^2 \leq \Sigma X_i^2 / \sigma^2 \leq C_2 \sigma_0^2 / \sigma^2}{\sigma^2} \right) = \int_{C_1 \sigma_0^2 / \sigma^2}^{C_2 \sigma_0^2 / \sigma^2} f_n(y) dy, \]

where \( f_n(y) \) is the pdf of \( X_n^2 \).

Fix \( C_1, C_2 \) by, for a size \( \alpha \) test,

\[ \int_{C_1}^{C_2} f_n(y) dy = 1 - \alpha, \]

and

\[ w'(\sigma_0^2) = 0. \]

Some manipulation reduces the second condition to \( C_{10}^{n/2} e^{-C_{10}^{n/2}} = C_{20}^{n/2} e^{-C_{20}^{n/2}}. \]

7.2 Let \( X_1, \ldots, X_n \) be an independent sample from \( N(\mu, \mu^2) \). Let \( T_1 = \bar{X} \) and \( T_2 = \sqrt{1/n} \sum X_i^2 \). Show that \( Z = T_1 / T_2 \) is ancillary. Explain why the Conditionality Principle would lead to inference about \( \mu \) being drawn from the conditional distribution of \( V = \sqrt{n}T_2 \) given \( Z \). Find the form of this conditional distribution.

[A7.2] Write \( U = \bar{X} \) and \( W = \sum_{i=1}^{n} (X_i - \bar{X})^2 / \mu^2 \). Then \( U \) and \( W \) are independent \( N(\mu, \mu^2/n) \) and \( \chi_{n-1}^2 \) respectively, so the joint density of \( (U, W) \) is

\[ f_{U,W}(u, w) = c_1 \mu^{-1} \exp\left(-\frac{n}{2\mu^2} (u - \mu)^2\right) w^{(n-3)/2} \exp\left(-\frac{w}{2}\right). \]

Here and below, \( c_1, c_2, \ldots \) denote generic constants, not depending on \( \mu \).

Transform to obtain the joint density of \( (V, Z) \). We have \( W = V^2 (1 - Z^2) / \mu^2, U = ZV / \sqrt{n}, \) and the Jacobian is \( 2V^2 / (\mu^2 \sqrt{n}) \), so the joint density of \( (V, Z) \) is

\[ c_2 \mu^{-1} \mu^{-(n-3)/2} \sqrt{n}^{-(3/2)} \mu^2 v^2 h_1(z) \exp\left(-\frac{n}{2\mu^2} (zv / \sqrt{n} - \mu)^2 - v^2 (1 - z^2) / (2\mu^2)\right), \]

where \( h_1(z) \) is some function of \( z \), not depending on \( \mu \).

Simplification shows this is

\[ c_3 h_2(z) \mu^{-n} v^{n-1} \exp\left(-\frac{1}{2} (v / \mu - z \sqrt{n})^2\right), \]
for some $h_2(z)$.

Observe that

$$
\int \mu^{-n}v^{n-1}\exp\left(-\frac{1}{2}(v/\mu - z\sqrt{n})^2\right)dv = h_3(z),
$$

say, not depending on $\mu$ (write $t = v/\mu$ and substitute), so we see that the marginal density of $Z$, obtained by integrating out $V$ from the joint density, does not depend on $\mu$; $Z$ is ancillary.

The minimal sufficient statistic $(T_1, T_2) \equiv (V, Z)$. Since $Z$ is ancillary, the Conditionality Principle implies that we should base inference on the conditional distribution of $V$ given $Z$. The conditional density is obtained by dividing the joint density of $V$ and $Z$ by the marginal density of $Z$, and is

$$
f(v \mid z; \mu) = c_4 \mu^{-n}v^{n-1}\exp\left(-\frac{1}{2}(v/\mu - z\sqrt{n})^2\right),
$$
directly from the above.

7.4 Suppose $X$ is normally distributed as $N(\theta, 1)$ or $N(\theta, 4)$, depending on whether the outcome, $Y$, of tossing a fair coin is heads ($y = 1$) or tails ($y = 0$). It is desired to test $H_0 : \theta = -1$ against $H_1 : \theta = 1$. Show that the most powerful (unconditional) size $\alpha = 0.05$ test is the test with rejection region given by $x \geq 0.598$ if $y = 1$ and $x \geq 2.392$ if $y = 0$.

Suppose instead that we condition on the outcome of the coin toss in construction of the tests. Verify that, given $y = 1$, the resulting most powerful size $\alpha = 0.05$ test would reject if $x \geq 0.645$ while, given $y = 0$ the rejection region would be $x \geq 2.290$.

[A7.4 Consider the situation where $X \mid I = 1$ is $N(\theta, \sigma_1^2)$ and $X \mid I = 2$ is $N(\theta, \sigma_2^2)$, with $P(I = 1) = P(I = 2) = 1/2$, and it is required to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 (> \theta_0)$. Use Neyman-Pearson to conclude that the most powerful test based on $(X, I)$ rejects $H_0$ when

$$
\frac{x - \frac{1}{2}(\theta_0 + \theta_1)}{2\sigma_i^2} \geq k,
$$

where $k$ is fixed to give the required size.

Check that for the case given $k = 0.299$: we require $\frac{1}{2}P(X \geq 2k \mid H_0, I = 1) + \frac{1}{2}P(X \geq 8k \mid H_0, I = 2) = 0.05$. Determination of the form of the conditional tests is trivial.

7.7 Let $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, q)$, with $X$ and $Y$ independent. Show that, as $p$ and $q$ range over $[0, 1]$, the joint distributions of $X$ and $Y$
form an exponential family. Show further that if \( p = q \) then

\[
P(X = x \mid X + Y = x + y) = \binom{m}{x} \binom{n}{y} / \binom{m + n}{x + y}.
\]

Hence find the form of a UMPU test of the null hypothesis \( H_0 : p \leq q \) against \( H_1 : p > q \).

In an experiment to test the efficacy of a new drug for treatment of stomach ulcers, 5 patients are given the new drug and 6 patients are given a control drug. Of the patients given the new drug, 4 report an improvement in their condition, while only 1 of the patients given the control drug reports improvement. Do these data suggest, at level \( \alpha = 0.1 \), that patients receiving the new drug are more likely to report improvement than patients receiving the control drug?

[This is again the hypergeometric distribution and the test presented here is conventionally referred to as Fisher's exact test for a 2 \times 2 table.]

[A7.7 The joint mass function]

\[
P\{X = x, Y = y\} = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} q^y (1-q)^{m-y}
\]

may be written in the form \( c(p,q) h(x,y) \exp(x\theta_1 + y\theta_2) \) where \( \theta_1 = \log(p/(1-p)), \quad \theta_2 = \log(q/(1-q)) \). The hypotheses being tested are equivalent to \( H_0 : \theta_1 \leq \theta_2 \) and \( H_1 : \theta_1 > \theta_2 \), since the joint mass function is proportional to \( \exp(x\lambda + (x + y)\theta_2) \), where \( \lambda = \theta_1 - \theta_2 \), and \( H_0 \equiv \lambda \leq 0 \), the UMPU test will be based on the conditional distribution of \( X \) given \( X + Y \). The conditional probability calculation is a simple manipulation: when \( p = q \), \( X + Y \) is Binomial(\( m + n \), \( p \)).

For the data given, calculate \( P(X = 4 \mid X + Y = 5) + P(X = 5 \mid X + Y = 5) \) < 0.1, so reject \( H_0 \) and conclude that improvement is more likely to be seen in experimental group.]