Imperial College London

MATH70090

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2023

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Bayesian Methods

Date: 2 May 2023

Time: 10:00 – 11:30 (BST)

Time Allowed: 1.5hrs

This paper has 3 Questions.

Please Answer Each Question in a Separate Answer Booklet.

Supplementary books may only be used after the relevant main books are full.

Any required additional material(s) will be provided.

Allow margins for marking.

Credit will be given for all questions attempted.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

The open-book material allowed during the examinations consists of any material provided by the lecturers and annotated by the students, i.e. annotated lecture notes, annotated slides, and annotated problem class sheets. Books and electronic devices are not allowed.

1. On a friend's computer the one visible line of code at the bottom of an `R` console is:

```
x <- bernoulli_trial(outputs = c(0, 1), probs = c(p, 1 - p))
```

which you assume performed a single Bernoulli trial with the two outcomes `0` and `1` having associated probabilities p and $1 - p$ respectively, and then stored the result as `x`. The value of p was presumably set earlier in the code, but with no access to this you have no reason to prefer any (valid) value of p to any other.

[In answering the below questions the result that, for integer $m, n \geq 0$,

$$\int_0^1 \mathrm{d}x \, x^m \, (1 - x)^n = \frac{m! \, n!}{(m + n + 1)!}$$

may be quoted without derivation.]

(i) Explain why your knowledge, K, at this point can be encoded probabilistically as a uniform distribution of the form $P(\mathtt{p} = p|K) = \mathrm{U}(p; 0, 1)$.

(ii) Given that the above line of code has already been executed, meaning that the value of `x` is now set, is it reasonable to assign a probability to the proposition that `x` is `0`? Briefly justify your answer.

(iii) Calculate the probability $P(\mathtt{x} = \mathtt{0}|K)$.

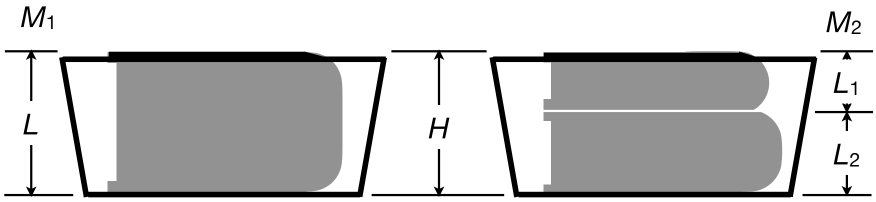
Scrolling back through the `R` window you see that the previous N calls to `bernoulli_trial` all resulted in the output `1`, information which is denoted as D.

(iv) Calculate the probability distribution $P(\mathtt{p} = p|D, K)$.

(v) Calculate the probability $P(\mathtt{x} = \mathtt{0}|D, K)$.

[Total 15 marks]

2. You are in the library looking for a notoriously thick textbook (e.g., *Bayesian Data Analysis*). There is a cart with books stacked horizontally, one of which just protrudes above the (opaque) side of the cart. There are two possible explanations:



- M_1: it is a single book of thickness L;
- M_2: it is a stack of two books of thicknesses L_1 and L_2.

(i) The only (prior) knowledge, K, you have about the books in the library is their mean thickness, $\bar{L}$. Use maximum entropy arguments to show that the implied state of knowledge about the thickness L of a book is

$$P(L|K) = \Theta(L)\,\frac{1}{\bar{L}}\,e^{-L/\bar{L}},$$

where $\Theta(x) = 1$ if $x > 0$ and $= 0$ otherwise is a (Heaviside) step function, which enforces the restriction that a book's thickness must be positive.

(ii) Explain why it might be reasonable to adopt "sharp" likelihoods of the form $P(H|L, M_1) = \delta_{\rm D}(H - L)$ and $P(H|L_1, L_2, M_2) = \delta_{\rm D}[H - (L_1 + L_2)]$, where $\delta_{\rm D}(\cdot)$ is the Dirac delta function. In particular, list any assumptions or approximations that are encoded in these expressions.

(iii) Using the above prior and likelihood, find the (marginal) likelihood $P(H|M_1, K)$.

(iv) Using the above prior and likelihood, find the (marginal) likelihood $P(H|M_2, K)$.

(v) Find an expression for the Bayes factor $B_{1,2}$ in favour of the single book hypothesis.
Calculate $B_{1,2}$ for the case that $H = 4\,\bar{L}$.

[Total 25 marks]

3. You are worried that your small pet is eating too much and so decide to weigh it, adopting a two-step process on your bathroom scales:

- You get on the scales by yourself and read off the measured weight, $\hat{W}_1$.
- You get on the scales holding the pet and read off the measured weight, $\hat{W}_2$.

If the scales were accurate then $\hat{W}_p = \hat{W}_2 - \hat{W}_1$ might be a reasonable point estimate of the pet's weight, W_p, although this ignores the fact that your own weight, W_o, is also unknown. Moreover, the user manual for the scales states that there is random measurement error with a standard deviation of $\sigma = 0.5$ kg, a source of uncertainty which should not be ignored.

(i) (a) Including the above information about the scales as part of your background knowledge K, explain why it is reasonable to model the measurement process using a normal sampling distribution of the form

$$P(\hat{W}|W, K) = \mathrm{N}(\hat{W}; W, \sigma^2),$$

where $\hat{W}$ is the measured weight and W is the true weight (of whatever is being weighed).

(b) The above sampling distribution could potentially yield a negative value for $\hat{W}$ if, e.g., $W \simeq \sigma$. Would this be a problem? Explain your reasoning.

(ii) You adopt the above sampling distribution and assume an improper flat prior distribution of the form $P(W_p, W_o|K) \propto \Theta(W_p)\,\Theta(W_o)$, where the Heaviside step functions, defined by $\Theta(x) = 1$ if $x > 0$ and $= 0$ otherwise, encode the fact that both weights are positive quantities.

(a) Find an expression for the joint posterior distribution $P(W_p, W_o|\hat{W}_1, \hat{W}_2, K)$. This does *not* need to be normalised, but it should be established whether it is normalisable.

(b) Find an expression for the marginal posterior distribution of your pet's weight, $P(W_p|\hat{W}_1, \hat{W}_2, K)$, that is valid under the assumption that $W_o \gg \sigma$. This again does *not* need to be normalised, but should be simplified as much as possible.

(c) Comment on this result in light of the initially suggested point estimate $\hat{W}_p = \hat{W}_2 - \hat{W}_1$.

(d) Sketch the marginal posterior distribution for W_p in the case that $\hat{W}_1 = 61.3$ kg and $\hat{W}_2 = 62.8$ kg.

(iii) A friend suggests the potentially annoying possibility that the scales are biased, in which case all measurements would be systematically offset by some constant but unknown amount, b.
Describe (e.g., in point form) how you would incorporate the possibility of such a bias into the above analysis.
How do you think the existence of such a bias would affect the final inference of your pet's weight? Explain your reasoning.

[Total 30 marks]

End of examination paper

Module: MATH70090
Setter: Daniel Mortlock
Checker: Nick Heard
Editor: Zak Varty
External: Dave Woods
Date: March 29, 2023

MSc EXAMINATIONS (STATISTICS)

MATH70090 Bayesian Methods
Time: 1 hour 30 minutes

Setter's signature	Checker's signature	Editor's signature
....	....	....

1. On a friend's computer the one visible line of code at the bottom of an `R` console is:

```
x <- bernoulli_trial(outputs = c(0, 1), probs = c(p, 1 - p))
```

which you assume performed a single Bernoulli trial with the two outcomes `0` and `1` having associated probabilities p and $1 - p$ respectively, and then stored the result as `x`. The value of p was presumably set earlier in the code, but with no access to this you have no reason to prefer any (valid) value of p to any other.

[In answering the below questions the result that, for integer $m, n \geq 0$,

$$\int_0^1 \mathrm{d}x\, x^m (1-x)^n = \frac{m!\, n!}{(m+n+1)!}$$

may be quoted without derivation.]

(i) Explain why your knowledge, K, at this point can be encoded probabilistically as a uniform distribution of the form $P(\mathrm{p} = p|K) = \mathrm{U}(p; 0, 1)$.

ANSWER: (SEEN)

As p is a probability it must be between 0 and 1.

The implied probability density must be constant in this range or otherwise at least some values p would be preferred, in contradicition with the stated assumptions.

Combining these two facts leads to the uniform distribution $P(\mathrm{p} = p|K) = \mathrm{U}(p; 0, 1)$. [2 marks]

(ii) Given that the above line of code has already been executed, meaning that the value of `x` is now set, is it reasonable to assign a probability to the proposition that `x` is `0`? Briefly justify your answer.

ANSWER: (SEEN)

Any definition of probability based on a state of knowledge or a degree of implication could reasonably be used to assign such a probability. (A definition based on randomness or long-run frequencies could not.)

[2 marks]

(iii) Calculate the probability $P(\mathtt{x} = \mathtt{0}|K)$.

ANSWER: (SEEN)

This probability is given by using the law of total probability to marginalise over the unknown value of p, which gives

$$P(\mathtt{x} = \mathtt{0}|K) = \int \mathrm{d}p\, P(p|K)\, P(\mathtt{x} = \mathtt{0}|p, K)$$

[This question continues on the next page . . .]

$$
\begin{aligned}
&= \int_0^1 \mathrm{d}p\, p \\
&= \left. \frac{p^2}{2} \right|_0^1 \\
&= \frac{1}{2}.
\end{aligned}
$$

[3 marks]

Scrolling back through the `R` window you see that the previous N calls to `bernoulli_trial` all resulted in the output `1`, information which is denoted as D.

(iv) Calculate the probability distribution $P(\mathrm{p} = p|D, K)$.

ANSWER: (SIMILAR TO SEEN)

The unnormalised posterior distribution for p is

$$
\begin{aligned}
P(\mathrm{p} = p|D, K) &\propto P(\mathrm{p} = p|K)\, P(D|\mathrm{p} = p, K) \\
&= \Theta(p)\, \Theta(1 - p)\, (1 - p)^N,
\end{aligned}
$$

where the independence of the trials motivates the use of a binomial distribution. This is normalised using the integral

$$
\begin{aligned}
\int \mathrm{d}p\, \Theta(p)\, \Theta(1 - p)\, (1 - p)^N &= \int_0^1 \mathrm{d}p\, (1 - p)^N \\
&= \left. -\frac{(1 - p)^{N+1}}{N + 1} \right|_0^1 \\
&= \frac{1}{N + 1},
\end{aligned}
$$

where the integral is given directly by the beta function definition. The final result is hence

$$P(\mathrm{p} = p|D, K) = \Theta(p)\, \Theta(1 - p)\, (N + 1)\, (1 - p)^N.$$

[4 marks]

(v) Calculate the probability $P(\mathtt{x} = \mathbb{0}|D, K)$.

ANSWER: (SIMILAR TO SEEN)

This is, again, given by using the law of total probability to marginalise over the unknown value of p, which this time gives

$$P(\mathtt{x} = \mathbb{0}|D, K) = \int \mathrm{d}p\, P(\mathrm{p} = p|D, K)\, P(\mathtt{x} = \mathbb{0}|\mathrm{p} = p, D, K)$$

[This question continues on the next page . . .]

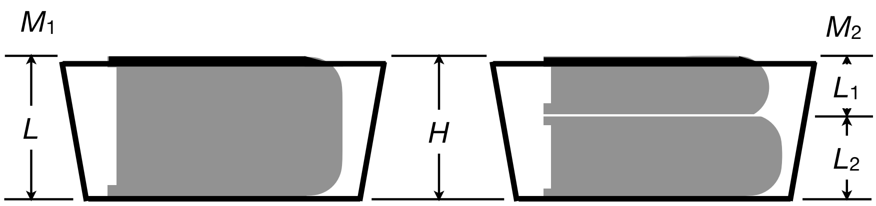
$$
\begin{aligned}
&= \int_0^1 \mathrm{d}p\,(N+1)\,(1-p)^N\,p \\
&= (N+1)\,\frac{1!\,N!}{(N+2)!} \\
&= \frac{1}{N+2},
\end{aligned}
$$

where the integral is evaluated directly using the beta function definition.

[4 marks]

[Total 15 marks]

2. You are in the library looking for a notoriously thick textbook (e.g., *Bayesian Data Analysis*). There is a cart with books stacked horizontally, one of which just protrudes above the (opaque) side of the cart. There are two possible explanations:



- M_1: it is a single book of thickness L;
- M_2: it is a stack of two books of thicknesses L_1 and L_2.

(i) The only (prior) knowledge, K, you have about the books in the library is their mean thickness, $\bar{L}$. Use maximum entropy arguments to show that the implied state of knowledge about the thickness L of a book is

$$P(L|K) = \Theta(L)\,\frac{1}{\bar{L}}\,e^{-L/\bar{L}},$$

where $\Theta(x) = 1$ if $x > 0$ and $= 0$ otherwise is a (Heaviside) step function, which enforces the restriction that a book's thickness must be positive.

ANSWER: (SIMILAR TO SEEN)

[The formalism here uses the notation from the lecture notes, which the students will have access to in the exam.]

Along with the normalisation requirement, the $J = 2$ available constraints on the form of the density $p(L) = P(L|K)$ are:

- $j = 0$: normalistion is $c_0 = 1$, with the constraint function is $f_0(L) = 1$;
- $j = 1$: mean is $c_1 = \bar{L}$, with the constraint function is $f_1(L) = L$;
- $j = 2$: $L > 0$, so $c_2 = 0$, with the constraint function is $f_2(L) = \Theta(-L)$.

For $j \in 0, 1, 2$ the relevant constraint can be expressed as

$$\int_{-\infty}^{\infty} \mathrm{d}L\, p(L)\, f_j(L) = c_j.$$

The maximum entropy distribution given these constraints has the form

$$\begin{aligned} p(L) &= c\,\exp\left[\sum_{j=1}^{1} \lambda_j\, f_j(L)\right] \\ &= c\,\exp[\lambda_1\, L + \lambda_2\,\Theta(-L)], \\ &= c\,e^{\lambda_1 L}\, e^{\Theta(-L)}, \end{aligned}$$

[This question continues on the next page . . .]

where c and the Lagrange multipliers λ_1 and λ_2 must be determined by the three constraints above.

The only way of satisfying the $j = 2$ constraint equation is to have $\lambda_2 \to -\infty$, so that $e^{\lambda_2 \Theta(-L)} \to \Theta(L)$, which is 0 if $L \leq 0$.

The normalisation requirement then gives c in terms of λ_1 as $c = \lambda_1$.

Finally, constraint on the mean then yields $\lambda_1 = -1/\bar{L}$.

The overall result is hence

$$P(L|K) = p(L) = \Theta(L)\,\frac{1}{\bar{L}}\,e^{-L/\bar{L}},$$

as required. [This could also be reached more simply by noting that positivity requirement means that the density $p(L)$ only has support for $L > 0$, so if this was applied directly then the second constraint could be omitted.]

[7 marks]

(ii) Explain why it might be reasonable to adopt "sharp" likelihoods of the form $P(H|L, M_1) = \delta_{\rm D}(H - L)$ and $P(H|L_1, L_2, M_2) = \delta_{\rm D}[H - (L_1 + L_2)]$, where $\delta_{\rm D}(\cdot)$ is the Dirac delta function. In particular, list any assumptions or approximations that are encoded in these expressions.

ANSWER: (UNSEEN)

The use of the delta function encodes the idea that the measurement uncertainty is negligible.

The absence of any offset between H and either L or $L_1 + L_2$ implies that the small amount by which the book(s) protrude over the edge of the cart can also be ignored, so effectively that the upper book cover is flush with the edge of the cart. [4 marks]

(iii) Using the above prior and likelihood, find the (marginal) likelihood $P(H|M_1, K)$.

ANSWER: (SEEN)

The (marginal) likelihood is given by the law of total probability as

$$\begin{aligned}
P(H|M_1, K) &= \int {\rm d}L\; P(H, L|M_1, K) \\
&= \int {\rm d}L\; P(L|M_1, K)\, P(H|L, M_1, K) \\
&= \int_0^\infty {\rm d}L\, \frac{1}{\bar{L}}\, e^{-L/\bar{L}}\, \delta_{\rm D}(H - L) \\
&= \frac{1}{\bar{L}}\, e^{-H/\bar{L}}.
\end{aligned}$$

[3 marks]

[This question continues on the next page . . .]

(iv) Using the above prior and likelihood, find the (marginal) likelihood $P(H|M_2, K)$.

ANSWER: (SIMILAR TO SEEN)

The (marginal) likelihood is given by the law of total probability as

$$\begin{aligned}
P(H|M_2, K) &= \int \mathrm{d}L_1 \int \mathrm{d}L_2 \, P(H, L_1, L_2|M_2, K) \\
&= \int \mathrm{d}L_1 \int \mathrm{d}L_2 \, P(L_1, L_2|M_2, K) \, P(H|L_1, L_2, M_2, K) \\
&= \int_0^\infty \mathrm{d}L_1 \int_0^\infty \mathrm{d}L_2 \, \frac{1}{\bar{L}} \, e^{-L_1/\bar{L}} \, \frac{1}{\bar{L}} \, e^{-L_2/\bar{L}} \, \delta_{\mathrm{D}}(H - L_1 - L_2) \\
&= \int_0^\infty \mathrm{d}L_1 \frac{1}{\bar{L}} \, e^{-L_1/\bar{L}} \int_0^\infty \mathrm{d}L_2 \, \frac{1}{\bar{L}} \, e^{-L_2/\bar{L}} \, \delta_{\mathrm{D}}[L_2 - (H - L_1)] \\
&= \int_0^\infty \mathrm{d}L_1 \, \frac{1}{\bar{L}^2} \, e^{-L_1/\bar{L}} \, \Theta(H - L_1) \, e^{-(H-L_1)/\bar{L}} \\
&= \frac{1}{\bar{L}^2} \, e^{-H/\bar{L}} \int_0^H \mathrm{d}L_1 \\
&= \frac{H}{\bar{L}^2} \, e^{-H/\bar{L}},
\end{aligned}$$

where the step function $\Theta(H - L_1)$ appears because the delta function is 0 for all positive L_2 if $L_1 > H$. (In qualitative terms, both L_1 and L_2 must both be less than H.) [8 marks]

(v) Find an expression for the Bayes factor $B_{1,2}$ in favour of the single book hypothesis.

Calculate $B_{1,2}$ for the case that $H = 4\,\bar{L}$.

ANSWER: (SIMILAR TO SEEN)

From above, the Bayes factor is

$$\begin{aligned}
B_{1,2} &= \frac{P(H|M_1, K)}{P(H|M_2, K)} \\
&= \frac{1/\bar{L} \, e^{-H/\bar{L}}}{H/\bar{L}^2 \, e^{-H/\bar{L}}} \\
&= \frac{\bar{L}}{H}
\end{aligned}$$

In the case that $H = 4\,\bar{L}$ the Bayes factor is hence $B_{1,2} = \bar{L}/(4\,\bar{L}) = 1/4$. [3 marks]

[Total 25 marks]

3. You are worried that your small pet is eating too much and so decide to weigh it, adopting a two-step process on your bathroom scales:

- You get on the scales by yourself and read off the measured weight, $\hat{W}_1$.
- You get on the scales holding the pet and read off the measured weight, $\hat{W}_2$.

If the scales were accurate then $\hat{W}_\mathrm{p} = \hat{W}_2 - \hat{W}_1$ might be a reasonable point estimate of the pet's weight, W_p, although this ignores the fact that your own weight, W_o, is also unknown. Moreover, the user manual for the scales states that there is random measurement error with a standard deviation of $\sigma = 0.5$ kg, a source of uncertainty which should not be ignored.

(i) (a) Including the above information about the scales as part of your background knowledge K, explain why it is reasonable to model the measurement process using a normal sampling distribution of the form

$$P(\hat{W}|W, K) = \mathrm{N}(\hat{W}; W, \sigma^2),$$

where $\hat{W}$ is the measured weight and W is the true weight (of whatever is being weighed).

ANSWER: (SEEN)
The stated knowledge, K, about the sampling distribution $P(\hat{W}|W)$ is only that it has variance σ^2. As there is no mention of any systematic bias, symmetry arguments imply that it has a mean of W. The maximum entropy distribution with mean W and variance σ^2 is $\mathrm{N}(\hat{W}; W, \sigma^2)$, so the maximum entropy principle leads to the suggested form of $P(\hat{W}|W, K)$.
[Other arguments are potentially reasonable, but some answers are definitely incorrect. In particular, it would be inappropriate to invoke the central limit theorem.]

[4 marks]

(b) The above sampling distribution could potentially yield a negative value for $\hat{W}$ if, e.g., $W \simeq \sigma$. Would this be a problem? Explain your reasoning.

ANSWER: (SIMILAR TO SEEN)
There is no problem, and in particular the possibility of inferring a negative weight from $\hat{W} < 0$ would be an artefact of the fundamentally arbitrary choice of an estimator.
Depending on the details of the measurement process (i.e., the form of the sampling distribution) the numerical datum associated with a positive-definite physical quantity, such as the weight of the pet, could easily be negative; this number has no physical meaning, and could be incorporated into any properly set up inference.

[2 marks]

[This question continues on the next page . . .]

(ii) You adopt the above sampling distribution and assume an improper flat prior distribution of the form $P(W_{\rm p}, W_{\rm o}|K) \propto \Theta(W_{\rm p})\,\Theta(W_{\rm o})$, where the Heaviside step functions, defined by $\Theta(x) = 1$ if $x > 0$ and $= 0$ otherwise, encode the fact that both weights are positive quantities.

(a) Find an expression for the joint posterior distribution $P(W_{\rm p}, W_{\rm o}|\hat{W}_1, \hat{W}_2, K)$. This does *not* need to be normalised, but it should be established whether it is normalisable.

ANSWER: (SEEN)
The unnormalised joint posterior distribution is

$P(W_{\rm p}, W_{\rm o}|\hat{W}_1, \hat{W}_2, K)$

$$\begin{aligned}
&\propto P(W_{\rm p}, W_{\rm o}|K)\,P(\hat{W}_1, \hat{W}_2|W_{\rm p}, W_{\rm o}, K)\\
&\propto \Theta(W_{\rm p})\,\Theta(W_{\rm o})\,{\rm N}(\hat{W}_1; W_{\rm o}, \sigma^2)\,{\rm N}(\hat{W}_2; W_{\rm o} + W_{\rm p}, \sigma^2)\\
&\propto \Theta(W_{\rm p})\,\Theta(W_{\rm o})\,\exp\left[-\frac{1}{2}\left(\frac{\hat{W}_1 - W_{\rm o}}{\sigma}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{\hat{W}_2 - (W_{\rm o} + W_{\rm p})}{\sigma}\right)^2\right]\\
&\propto \Theta(W_{\rm p})\,\Theta(W_{\rm o})\,{\rm N}\left(W_{\rm o}; \frac{\hat{W}_1 + \hat{W}_2 - W_{\rm p}}{2}, \frac{\sigma^2}{2}\right){\rm N}(W_{\rm p}; \hat{W}_2 - \hat{W}_1, 2\,\sigma^2).
\end{aligned}$$

A number of different algebraic forms are possible, although full marks only for some level of simplification.
This Gaussian form for the posterior given above shows that this is normaliseable, and hence proper, as the density decreases (more than) exponentially as both $W_{\rm o} \to \infty$ and $W_{\rm p} \to \infty$. [4 marks]

(b) Find an expression for the marginal posterior distribution of your pet's weight, $P(W_{\rm p}|\hat{W}_1, \hat{W}_2, K)$, that is valid under the assumption that $W_{\rm o} \gg \sigma$. This again does *not* need to be normalised, but should be simplified as much as possible.

ANSWER: (SIMILAR TO SEEN)
The unnormalised marginal posterior is given by integrating the unnormalised joint posterior over $W_{\rm o}$ to give

$P(W_{\rm p}|\hat{W}_1, \hat{W}_2, K)$

$$\begin{aligned}
&\propto \int {\rm d}W_{\rm o}\,P(W_{\rm p}, W_{\rm o}|\hat{W}_1, \hat{W}_2, K)\\
&= \int_0^\infty {\rm d}W_{\rm o}\,\Theta(W_{\rm p})\,{\rm N}\left(W_{\rm o}; \frac{\hat{W}_1 + \hat{W}_2 - W_{\rm p}}{2}, \frac{\sigma^2}{2}\right){\rm N}(W_{\rm p}; \hat{W}_2 - \hat{W}_1, 2\,\sigma^2)
\end{aligned}$$

[This question continues on the next page . . .]

$$\begin{aligned}
&\simeq \Theta(W_p)\, \mathrm{N}(W_p; \hat{W}_2 - \hat{W}_1, 2\,\sigma^2) \int_{-\infty}^{\infty} \mathrm{d}W_o\, \mathrm{N}\left(W_o; \frac{\hat{W}_1 + \hat{W}_2 - W_p}{2}, \frac{\sigma^2}{2}\right) \\
&= \Theta(W_p)\, \mathrm{N}(W_p; \hat{W}_2 - \hat{W}_1, 2\,\sigma^2)
\end{aligned}$$

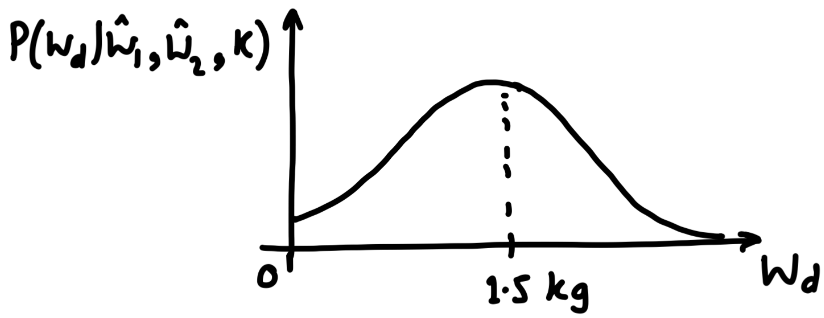
where the third step invokes the approximation that $W_o \gg \sigma$. [7 marks]

(c) Comment on this result in light of the initially suggested point estimate $\hat{W}_p = \hat{W}_2 - \hat{W}_1$.

ANSWER: (SIMILAR TO SEEN)
If $\hat{W}_2 - \hat{W}_1 > 0$ then the peak/mode of the posterior distribution is at $W = \hat{W}_2 - \hat{W}_1$, recovering the initial point estimate, but now with an associated uncertainty. But if $\hat{W}_2 - \hat{W}_1 \leq 0$ then $\hat{W}_p$ would be negative, a nonsenical result, whereas the posterior distribution, while peaked at $W_p = 0$, would have support only for positive values. [3 marks]

(d) Sketch the marginal posterior distribution for W_p in the case that $\hat{W}_1$ = 61.3 kg and $\hat{W}_2$ = 62.8 kg.

ANSWER: (SIMILAR TO SEEN)
The posterior distribution in this case has the form:



[3 marks]

(iii) A friend suggests the potentially annoying possibility that the scales are biased, in which case all measurements would be systematically offset by some constant but unknown amount, b.

Describe (e.g., in point form) how you would incorporate the possibility of such a bias into the above analysis.

How you think the existence of such a bias would affect the final inference of your pet's weight? Explain your reasoning.

ANSWER: (UNSEEN)

The bias, b would have to be treated as an unknown nuisance parameter to be included in the inference and then marginalised out. The main steps of such a

[This question continues on the next page . . .]

calculation would be (with K' distinct from K because of the information about the bias):

- Adopt a (presumably uninformative) prior distribution $P(b|K')$.
- Find the joint posterior distribution $P(W_\mathrm{p}, W_\mathrm{o}, b|\hat{W}_1, \hat{W}_2, K')$, where the likelihood now has the modified form $P(\hat{W}|W, K') = \mathrm{N}(\hat{W}; W + b, \sigma^2)$.
- Integrate out W_o (as above) and b to find the marginalised posterior distribution $P(\hat{W}|\hat{W}_1, \hat{W}_2, K')$.

Given that b affects both measurements in the same way, it could be expected to largely cancel out, and this would be the case if $|b| \ll W_\mathrm{o}$ or, equivalently, that $|b| \ll \hat{W}_1$. However there will not be a formal cancellation, as could be seen if $b \simeq -W_\mathrm{o}$, in which case the prior that the true weights cannot be negative would couple the posterior knowledge of the three parameters. [7 marks]

[Total 30 marks]