Statistical Data Analysis 2025/26
Exam Revision Session

London Postgraduate Lectures on Particle Physics

University of London MSc/MSci course PH4515

Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

Course web page via RHUL moodle (PH4515) and also

www.pp.rhul.ac.uk/~cowan/stat_course.html

General Info

The exam is in-person. Full details will have been communicated to you by RHUL administration. If not, please contact epms-school@rhul.ac.uk

The exam time is 2 ½ hours, with extra time for special circumstances authorised as appropriate.

Exam counts 80%, coursework 20% of module; module pass mark is 50.

The examination is closed-book and no notes are allowed.

You do not need a calculator for the exam.

The final details regarding venue, times, special arrangements, etc., are made through the RHUL admin team in conjunction with your College's Physics Department – please turn to them (not me) for any questions about these matters.

Every effort has been made to ensure that the exam does not contain any errors. If you believe a question has a typo, it will not be possible to query this during the exam. If you suspect a question contains an error, you should explain what you believe the problem is and proceed on the basis of your stated assumptions.

Exam format

Answer any 3 questions out of the 5.

No credit is given for work on any further questions.

You must indicate on the cover of the exam book which questions you have attempted and want to have considered for marking.

Just to be safe, if you have a partially attempted question in your exam booklet that you do not want to have marked, it is best to indicate this clearly in the booklet.

If you write down answers for more than 3 questions but do not indicate which you want to have considered, then you run the risk that the questions marked be not those that you intended.

Reminder of revision resources

There are past exams and (some) solutions on:

http://www.pp.rhul.ac.uk/~cowan/ph4515/

Solutions for the papers since 2011 will not be posted, except for those problems worked in discussion sessions.

Remember that the solutions to the problem sheets are contained in the discussion notes on moodle (PS *n* discussed in week *n*+3).

And here is a note with some advice on reviewing for the exam:

http://www.pp.rhul.ac.uk/~cowan/ph4515/ph4515_review.pdf

You should be able to make relevant approximations where appropriate (e.g., when prompted to do so in the question), using, for example, expansions for small ε such as

$$\begin{aligned} \ln(1+\varepsilon) &\approx \varepsilon \,, \\ (1+\varepsilon)^a &\approx 1 + a\varepsilon \,. \end{aligned}$$

Brief course overview

1 Probability, Bayes' theorem
2 Random variables and probability densities
3 Expectation values, error propagation
4 Catalogue of pdfs
5 The Monte Carlo method
6 Statistical tests: general concepts
7 Test statistics, multivariate methods
8 Goodness-of-fit tests
9 Parameter estimation, maximum likelihood
10 More maximum likelihood
11 Method of least squares
12 Interval estimation, setting limits
13 Nuisance parameters, systematic uncertainties
14 Examples of Bayesian approach

Probability, Bayes' theorem

Kolmogorov axioms

$$\boxed{\begin{array}{l} \text{For all } A \subset S, P(A) \geq 0 \\ \qquad P(S) = 1 \\ \text{If } A \cap B = \emptyset, P(A \cup B) = P(A) + P(B) \end{array}}$$

Conditional probability

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

Bayes' theorem

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}}$$

Interpretations: frequentist, subjective (degree of belief, Bayesian)

Random variables and probability densities

Discrete → probability (mass) function

$$P(x_i) = p_i \qquad \sum_i P(x_i) = 1$$

Continuous → probability density function (pdf)

$$\int_a^b f(x)\,dx = P(a \le x \le b)$$

Cumulative distribution

$$\int_{-\infty}^{x} f(x')\,dx' \equiv F(x)$$

Joint pdf

$$\int \cdots \int_R f(x_1, \ldots, x_n)\,dx_1 \cdots dx_n = P(\mathbf{x} \in R)$$

Marginal pdf

$$f_x(x) = \int f(x, y)\,dy$$

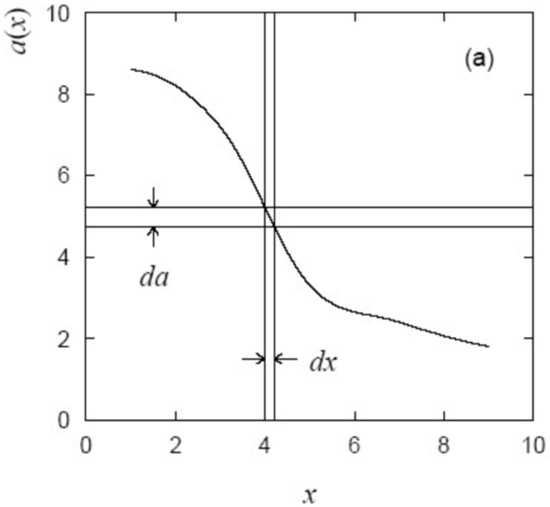
Conditional pdf

$$h(y|x) = \frac{f(x,y)}{f_x(x)} \qquad g(x|y) = \frac{f(x,y)}{f_y(y)}$$

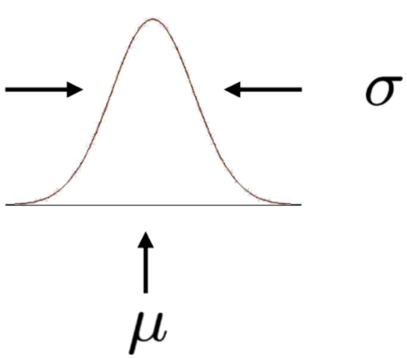
Functions of r.v.s, expectation values, error propagation

Function $a(\boldsymbol{x})$ of random variable $\boldsymbol{x}$,
pdf of a:

$$g(a)\,da = \int_{dS} f(x)\,dx$$



Expectation (mean) of x $\quad E[x] = \int x f(x)\,dx \equiv \mu$



Variance of x $\quad V[x] = E[x^2] - \mu^2 = E[(x-\mu)^2]$

Covariance of x, y $\quad \mathrm{cov}[x,y] = E[xy] - \mu_x \mu_y$

"Error propagation" for $V[y(\boldsymbol{x})]$: $\quad V[y] \approx \sum_{i,j} \left(\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right)_{\mathbf{x}=\boldsymbol{\mu}} \mathrm{cov}[x_i, x_j]$

Catalogue of pdfs

and many more
(but know these)

Binomial $$P(m;N,p) = \frac{N!}{m!(N-m)!} p^m (1-p)^{N-m} \,, \quad E[m] = Np \,,$$
$$V[m] = Np(1-p)$$

Poisson $$P(n|\nu) = \frac{\nu^n}{n!} e^{-\nu} \,, \quad n = 0, 1, \ldots \,, \quad E[n] = \nu \,, \quad V[n] = \nu$$

Uniform $$f(x|\alpha,\beta) = \frac{1}{\beta - \alpha} \,, \quad \alpha \le x \le \beta \,, \quad E[x] = \frac{\alpha+\beta}{2} \,, \quad V[x] = \frac{\beta - \alpha}{12}$$

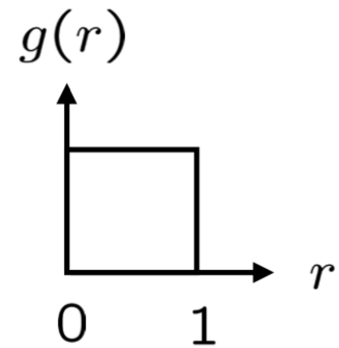
Exponential $$f(x|\xi) = \frac{1}{\xi} e^{-x/\xi} \,, \quad 0 \le x < \infty \,, \quad E[x] = \xi \,, \quad V[x] = \xi^2$$

Gaussian $$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \,, \quad E[x] = \mu \,, \quad V[x] = \sigma^2$$

Central Limit Theorem: distribution of $y = \sum_{i=1}^{n} x_i$

tends to Gaussian when $n \to \infty$ (under specific conditions)

The Monte Carlo method



Random number generator to produce $r \sim \mathrm{U}[0,1]$

$$\rightarrow r_1, \ldots, r_n$$

Convert sequence of r values to x values for $x \sim f(x)$

Transformation method:

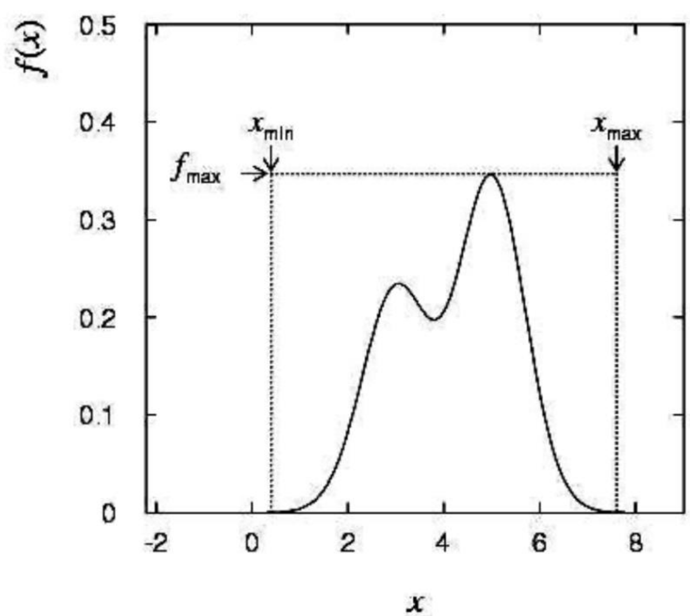
Find cumulative distribution $F(x)$

Set $F(x) = r$ and solve for $x(r)$

Acceptance-rejection method:

Generate points uniform under enclosing curve

Accept x if point is below $f(x)$.



Hypothesis tests: general concepts

Define test of H_0 by specifying critical region W:

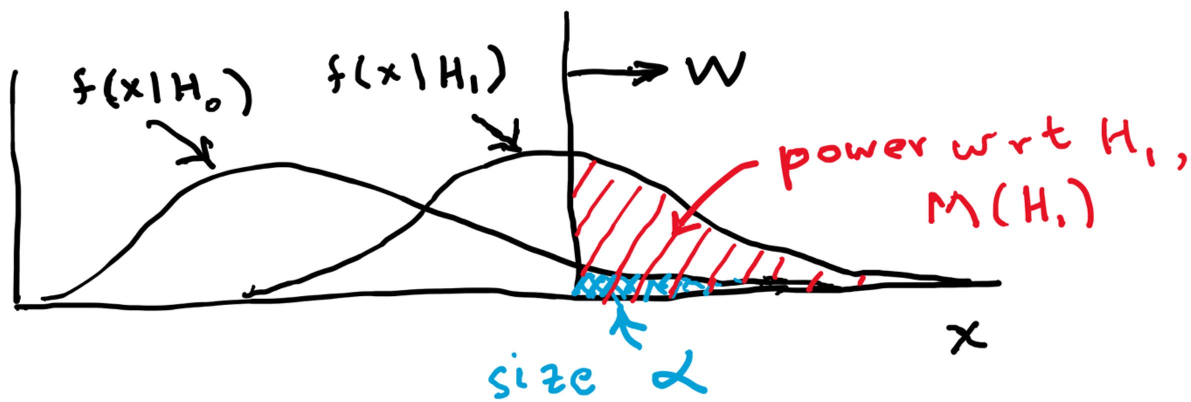
$$P(x \in W \mid H_0) \le \alpha$$

α = size of test (small, e.g., 0.05)

If observe data x in W, reject H_0.

Choose critical region to maximize power with respect to relevant alternative H_1:

$$M(H_1) = P(x \in W | H_1)$$



Test statistics, multivariate methods

Define boundary of critical region by statistic $t(\boldsymbol{x}) = t_{\mathrm{c}}$.

Neyman-Pearson: optimal *W* from $$t(\mathbf{x}) = \frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)}$$

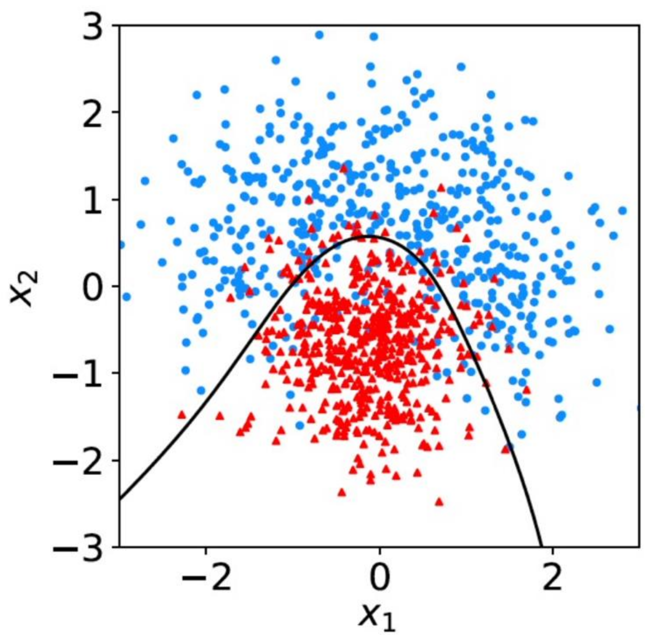
But often cannot compute likelihood ratio, need to use multivariate methods (Machine Learning), adjust parameters of statistic using training data.

Linear discriminant (Fisher)

Neural networks

Naive Bayes, KDE

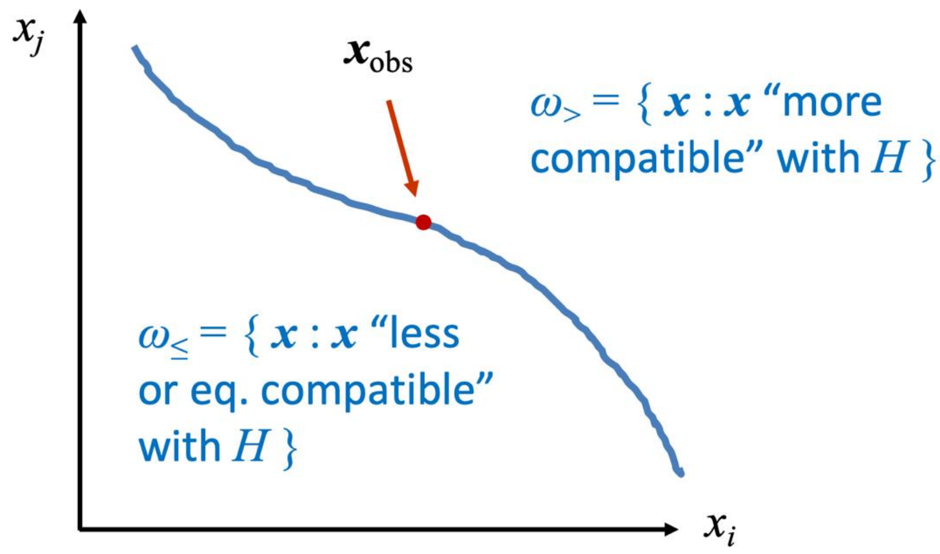
Boosted Decision Trees



Issues: overtraining

Goodness-of-fit tests

Quantify compatibility between observation x_{obs} and predictions of hypothesis $P(x|H)$ with p-value:

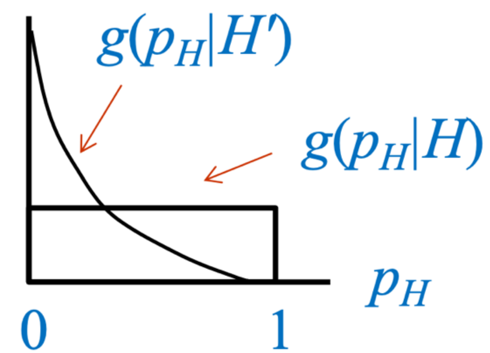


$$p = P(\mathbf{x} \in \omega_{\leq}(\mathbf{x}_{\text{obs}})|H)$$

= probability, under assumption of H, to observe data as discrepant with H as the data we got or more so.

Define region of equal/lesser compatibility with $t(\boldsymbol{x}) \geq t_{\text{obs}}$

$$p_H = \int_{t_{\text{obs}}}^{\infty} f(t|H)\, dt = \int_{\{\mathbf{x}\,:\,t(\mathbf{x}) \geq t_{\text{obs}}\}} f(\mathbf{x}|H)\, d\mathbf{x}$$



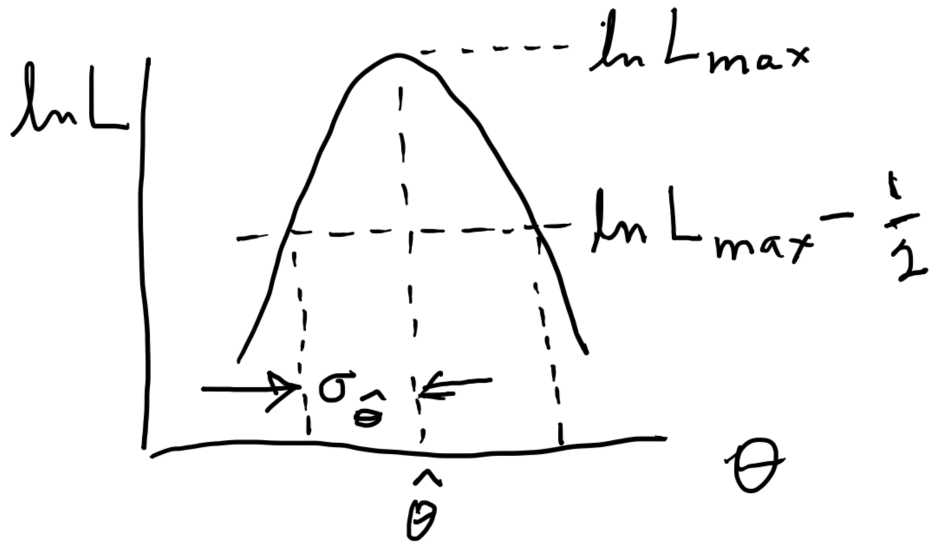
Parameter estimation, maximum likelihood

$\widehat{\theta}(\vec{x})$ = function of data (written with hat)

Want small bias $b = E[\widehat{\theta}] - \theta$ and small variance $V[\widehat{\theta}]$

(cannot in general minimize both simultaneously)

Maximum Likelihood:
maximize $L(\theta) = P(\boldsymbol{x}|\theta)$



Variance of MLE:

- Sometimes closed form
- Monte Carlo
- Information inequality

$$V[\widehat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 \Big/ E\left[-\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

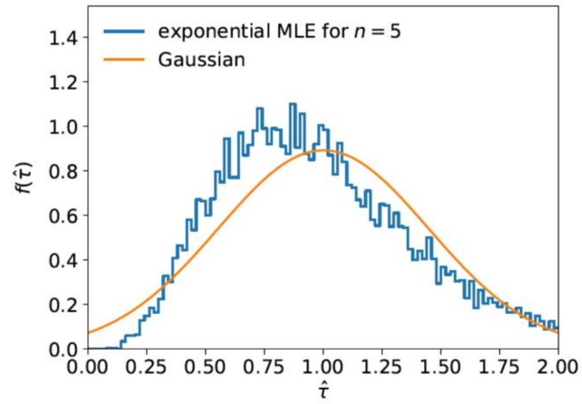
- Graphical method

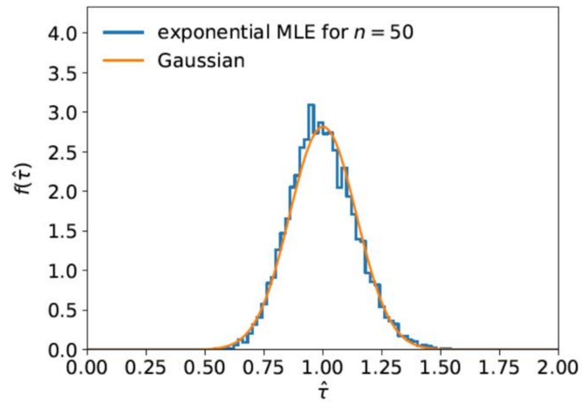
Asymptotic properties of likelihood and MLE

Likelihood becomes Gaussian in shape

MLE consistent, asymptotically unbiased

MLE Gaussian distributed





Information inequality becomes equality:

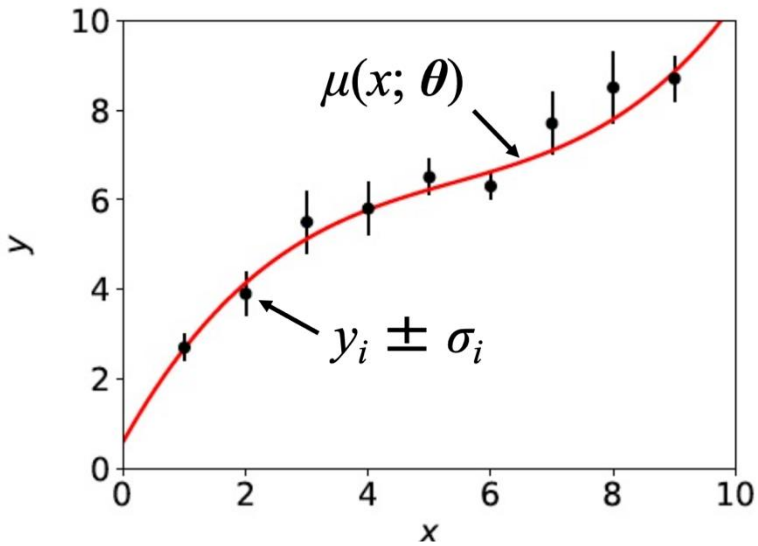
$$V_{ij}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right]$$

Estimate using $\widehat{V}_{ij}^{-1} = -\left.\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right|_{\hat{\boldsymbol{\theta}}}$

Method of least squares

Independent measurements:

$y_i \sim \text{Gauss}(\mu(x_i,\theta),\ \sigma_i),\ i=1,\ldots,N$



Maximizing $\ln L$ equivalent to minimizing

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \mu(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2} = -2\ln L(\boldsymbol{\theta}) + \text{const.}$$

or for correlated data $\quad \chi^2(\boldsymbol{\theta}) = (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T V^{-1} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))$

For variances of estimators use same technology as for MLE.

Minimized $\chi^2 \rightarrow p$-value (goodness of fit)

Interval estimation, setting limits

Test (size α) hypothesized parameter value θ for all θ.

Set of θ values not rejected = confidence interval at CL = $1 - \alpha$

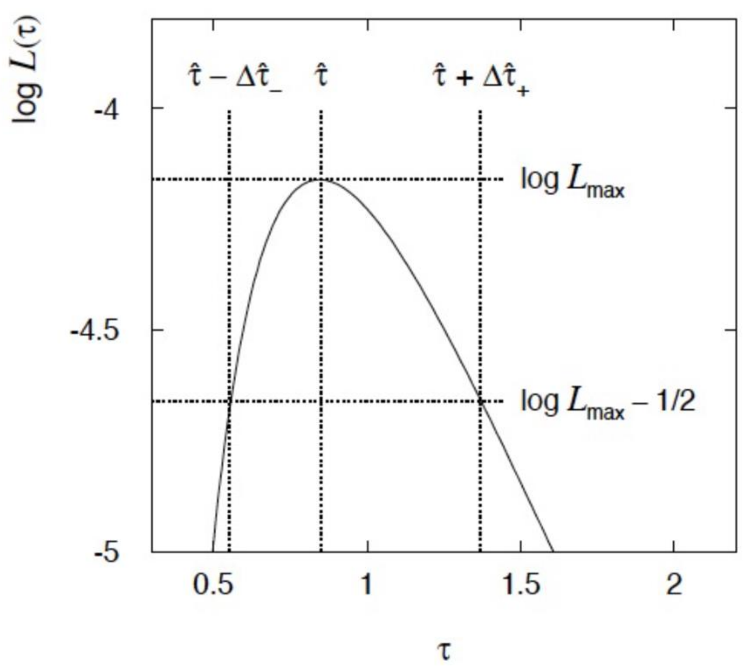
Equivalently, critical region of test = data that gives $p_\theta \leq \alpha$

To find boundary of interval, set $p_\theta = \alpha$ and solve for θ.

P(conf. interval "covers" $\theta | \theta$) $\geq 1 - \alpha$

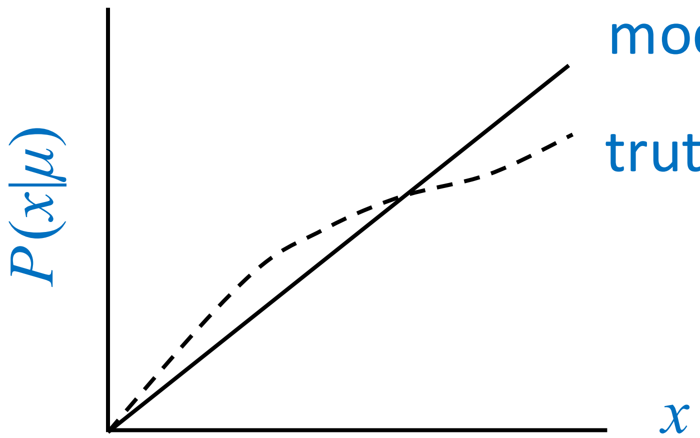
Approximate construction from likelihood function

$$\ln L(\boldsymbol{\theta}) = \ln L(\hat{\boldsymbol{\theta}}) - \tfrac{1}{2} F^{-1}_{\chi^2_n}(1 - \alpha)$$



Systematic uncertainties and nuisance parameters

In general, our model of the data is not perfect:



model: $P(x|\mu) = \mu x$

truth: $P(x|\mu) = \mu x + \alpha x^2 + \beta x^3 + \cdots$

Can improve model by including additional adjustable parameters.

$$P(x|\mu) \rightarrow P(x|\mu, \boldsymbol{\theta})$$

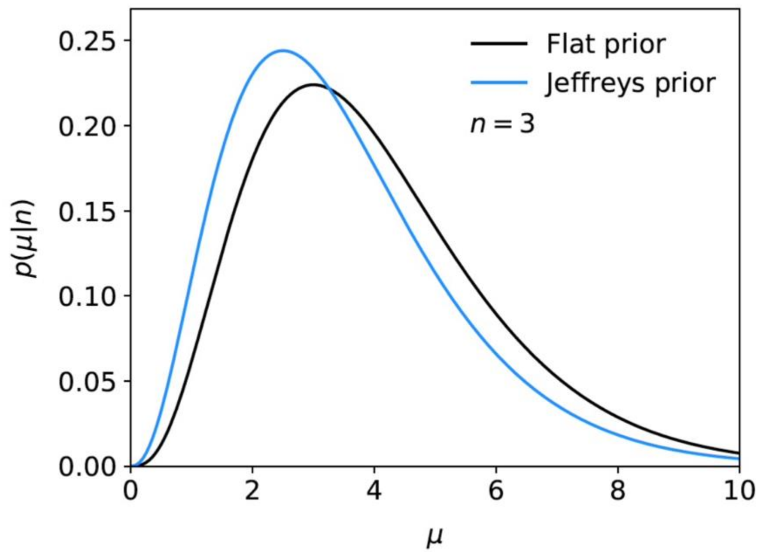
Nuisance parameter ↔ systematic uncertainty. Some point in the parameter space of the enlarged model should be “true”.

Presence of nuisance parameter decreases sensitivity of analysis to the parameter of interest (e.g., increases variance of estimate).

Some Bayesian methods

Bayes' theorem gives

$$p(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int L(x|\theta')\pi(\theta')\,d\theta'}$$



Bayesian interval/limit from e.g.

$$1 - \alpha = \int_0^{\mu_{\rm up}} p(\mu|n)\,d\mu$$

MCMC for marginalizing posterior (Metropolis-Hastings)

$$p(\theta_0|x) = \int p(\theta_0, \theta_1|x)\,d\theta_1$$

Bayes factor for model selection:

$$B_{ij} = \frac{P(\mathbf{x}|H_i)}{P(\mathbf{x}|H_j)}$$

= posterior odds if one takes prior odds equal to one.

2019 Exam Question 5(a)

5. Suppose y_i follows a Gaussian distribution with unknown mean μ and known standard deviations σ_i, and one has an independent sample $\vec{y} = (y_1, \ldots, y_N)$.

   (a) Write down the likelihood function for μ and show that the maximum-likelihood estimator is

$$\hat{\mu} = \frac{\sum_{i=1}^N y_i/\sigma_i^2}{\sum_{i=1}^N 1/\sigma_i^2} \; .$$

[6]

5(a) [6 marks] The N Gaussian variables y_i are independent and so the likelihood function is the product of their pdfs,

$$L(\mu) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i - \mu)^2/2\sigma_i^2} \; ,$$

and the log-likelihood is therefore

$$\ln L(\mu) = -\frac{1}{2} \sum_{i=1}^N \frac{(y_i - \mu)^2}{\sigma_i^2} + C \; ,$$

where C represents terms that do not depend on μ.

2019 Exam Question 5(a)

5. Suppose y_i follows a Gaussian distribution with unknown mean μ and known standard deviations σ_i, and one has an independent sample $\vec{y} = (y_1, \ldots, y_N)$.

   (a) Write down the likelihood function for μ and show that the maximum-likelihood estimator is

$$\hat{\mu} = \frac{\sum_{i=1}^{N} y_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2} \ .$$

[6]

5(a) [6 marks] The N Gaussian variables y_i are independent and so the likelihood function is the product of their pdfs,

$$L(\mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i-\mu)^2/2\sigma_i^2} \ ,$$

and the log-likelihood is therefore

$$\ln L(\mu) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i-\mu)^2}{\sigma_i^2} + C \ ,$$

where C represents terms that do not depend on μ.

2019 Exam Question 5(a)

Setting the derivative of $\ln L$ to zero,

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{N} \frac{(y_i - \mu)}{\sigma_i^2} = 0 \; ,$$

and solving for μ gives the maximum-likelihood estimator

$$\hat{\mu} = \frac{\sum_{i=1}^{N} y_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2} \; .$$

2019 Exam Question 5(b)

(b) Show that the estimator is unbiased and find its variance. [6]

5(b) [6 marks] Using $E[y_i] = \mu$, the expectation value of $\hat{\mu}$ is

$$E[\hat{\mu}] = E\left[\frac{\sum_{i=1}^N y_i/\sigma_i^2}{\sum_{i=1}^N 1/\sigma_i^2}\right] = \frac{1}{\sum_{i=1}^N 1/\sigma_i^2} \sum_{i=1}^N E[y_i]/\sigma_i^2 = \mu \ ,$$

and therefore the estimator is unbiased. Using $V[y_i] = \sigma_i^2$, the variance is found to be

$$V[\hat{\mu}] = V\left[\frac{\sum_{i=1}^N y_i/\sigma_i^2}{\sum_{i=1}^N 1/\sigma_i^2}\right] = \frac{1}{\left(\sum_{i=1}^N 1/\sigma_i^2\right)^2} \sum_{i=1}^N V[y_i]/\sigma_i^4 = \frac{1}{\sum_{i=1}^N 1/\sigma_i^2} \ .$$

2019 Exam Question 5(c)

(c) Explain how the method of least squares stands in relation to the method of maximum likelihood for this problem. [2]

5(c) [2 marks] As can be seen from the expression for $\ln L$ from (a), maximizing the log-likelihood function is equivalent to minimizing

$$\chi^2(\mu) = \sum_{i=1}^{N} \frac{(y_i - \mu)^2}{\sigma_i^2} ,$$

and so for this problem the methods of maximum likelihood and least squares are equivalent.

2019 Exam Question 5(d)

For parts (d) and (e) consider the Bayesian approach to inference about μ.

(d) Show that the Jeffreys prior $\pi_{\mathrm{J}}(\mu)$ is a constant. [6]

5(d) [6 marks] The Jeffreys prior is $\pi_{\mathrm{J}} \propto \sqrt{I}$, where $I = -E[\partial^2 \ln L / \partial\mu^2]$ is the Fisher information. The second derivative of $\ln L$ is

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\sum_{i=1}^{N} \frac{1}{\sigma_i^2}$$

which is independent of the data and thus equal to its expectation value. The Fisher information is therefore

$$I = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \, .$$

Therefore the Jeffreys prior is

$$\pi_{\mathrm{J}}(\mu) \propto \sqrt{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \, ,$$

which is constant as it does not depend on μ.

2019 Exam Question 5(e)

(e) Using the Jeffreys prior, show that the posterior probability $p(\mu|\vec{y})$ has the form

$$p(\mu|\vec{y}) \propto \exp\left[-\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\sigma_{\hat{\mu}}^2}\right] ,$$

where $\hat{\mu}$ and $\sigma_{\hat{\mu}}$ are the same as the maximum-likelihood estimator and its standard deviation, as found in (a) and (b). (Hint: for the argument of the exponential, assume a quadratic function in μ and relate the coefficients to those of a Taylor series, then complete the square.) [12]

5(e) [12 marks] Bayes' theorem says that the posterior for μ is

$$p(\mu|\vec{y}) \propto p(\vec{y}|\mu)\pi_{\rm J}(\mu) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\frac{(y_i-\mu)^2}{\sigma_i^2}\right] \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i-\mu)^2}{\sigma_i^2}\right] .$$

The argument of the exponential is a quadratic function in μ, so we can write it as

$$f(\mu) = \sum_{i=1}^{N}\frac{(y_i-\mu)^2}{\sigma_i^2} = a\mu^2 + b\mu + c = f(0) + \left.\frac{\partial f}{\partial \mu}\right|_{\mu=0} \mu + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial \mu^2}\right|_{\mu=0} \mu^2$$

2019 Exam Question 5(e)

By identifying a, b and c as the coefficients of a Taylor series, we have

$$
\begin{aligned}
c &= f(0) = \sum_{i=1}^{N} \frac{y_i^2}{\sigma_i^2}\,, \\
b &= \left.\frac{\partial f}{\partial \mu}\right|_{\mu=0} = -2\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}\,, \\
a &= \frac{1}{2!}\left.\frac{\partial^2 f}{\partial \mu^2}\right|_{\mu=0} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\,.
\end{aligned}
$$

To relate this to the given quadratic function in μ we complete the square,

$$
f(\mu) = a\left(\mu^2 + \frac{b}{a}\mu + \frac{c}{a}\right) = a\left(\mu^2 + \frac{b}{a} + \frac{b^2}{4a^2} - \frac{b^2}{4a^2} + \frac{c}{a}\right) = a\left(\mu + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c\,,
$$

2019 Exam Question 5(e)

and therefore we can write

$$\sum_{i=1}^{N} \frac{(y_i - \mu)^2}{\sigma_i^2} = \frac{(\mu - \hat{\mu})^2}{\sigma_{\hat{\mu}}^2} + C$$

with

$$\hat{\mu} = -\frac{b}{2a} = \frac{\sum_{i=1}^{N} y_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2} \,, \qquad \sigma_{\hat{\mu}}^2 = \frac{1}{a} = 1 \Bigg/ \sum_{i=1}^{N} 1/\sigma_i^2 \,, \qquad C = -\frac{b^2}{4a} + c \,,$$

which shows the desired result where $\hat{\mu}$ and $\sigma_{\hat{\mu}}^2$ are indeed the same as found for the maximum-likelihood estimator.

2019 Exam Question 5(f)

For the following parts suppose the result of the observation is characterized only by $\hat{\mu}$, which is Gaussian distributed about μ with standard deviation $\sigma_{\hat{\mu}}$. We wish to test the hypothesis $\mu = \mu_0$ with respect to an alternative $\mu = \mu_1 < \mu_0$.

(f) Sketch the distribution of $\hat{\mu}$ and show a critical region for the test, indicating its size α and power M with respect to the alternative. Show that the critical region of the test is given by $\hat{\mu} \le \mu_c$, where

$$\mu_c = \mu_0 - \sigma_{\hat{\mu}} \Phi^{-1}(1 - \alpha) \ ,$$

where Φ^{-1} is the quantile of the standard Gaussian. [6]

5(f) [6 marks] A sketch of the distribution of of $\hat{\mu}$ for $\mu = \mu_0$ and $\mu = \mu_1 < \mu_0$ is shown in Fig. 4.

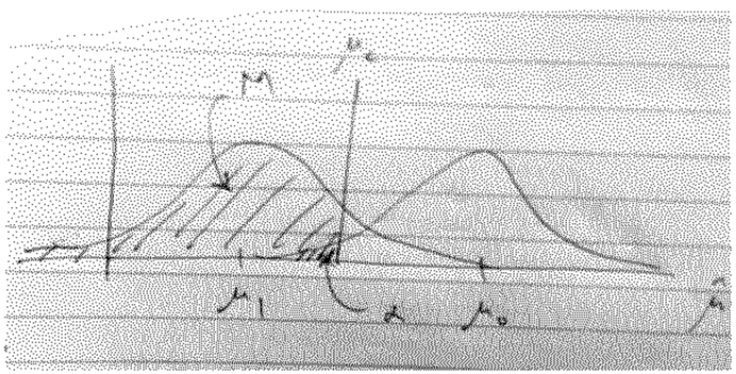


Figure 4: Distributions of $\hat{\mu}$ for $\mu = \mu_0$ and $\mu = \mu_1 < \mu_0$, indicating the critical region $\hat{\mu} < \mu c$, the size α and power M of the test.

2019 Exam Question 5(f)

The boundary of the critical region μ_c for a test of size α of $\mu = \mu_0$ is determined by the requirement

$$\alpha = P(\hat{\mu} < \mu_c | \mu_0) = \int_{-\infty}^{\mu_c} \frac{1}{\sqrt{2\pi}\sigma_{\hat{\mu}}} e^{-(\hat{\mu}-\mu_0)^2/2\sigma_{\hat{\mu}}^2} \, d\hat{\mu} = \Phi\left(\frac{\mu_c - \mu_0}{\sigma_{\hat{\mu}}}\right) ,$$

where Φ is the standard Gaussian cumulative distribution. Solving for μ_c gives

$$\mu_c = \mu_0 + \sigma_{\hat{\mu}} \Phi^{-1}(\alpha) = \mu - \sigma_{\hat{\mu}} \Phi^{-1}(1 - \alpha) ,$$

where for the final equality we used the property of the standard Gaussian quantile $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$.

2019 Exam Question 5(g)

(g) Find the power of the test M with respect to the alternative $\mu = \mu_1$ in terms of μ_c and α. [2]

5(g) [2 marks] The power of the test relative the alternative $\mu = \mu_1$ is

$$M_1 = P(\hat{\mu} < \mu_c | \mu_1) = \int_{-\infty}^{\mu_c} \frac{1}{\sqrt{2\pi}\sigma_{\hat{\mu}}} e^{-(\hat{\mu}-\mu_1)^2/2\sigma_{\hat{\mu}}^2} \, d\hat{\mu} = \Phi\left(\frac{\mu_c - \mu_1}{\sigma_{\hat{\mu}}}\right) .$$

2019 Exam Question 3(a)

3. Suppose that the outcome of a measurement consists of two independent values, y and v, which follow

$$f_y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} ,$$

$$f_v(v) = \frac{\left(\frac{\nu}{2\sigma^2}\right)^{\nu/2}}{\Gamma(\nu/2)} v^{\nu/2-1} e^{-\nu v/2\sigma^2} .$$

Here Γ is the Euler gamma function and ν is a known constant. Suppose the parameters μ and σ^2 are both unknown; μ is the parameter of interest and σ^2 is a nuisance parameter.

(a) Show that the log-likelihood function can be written as

$$\ln L(\mu, \sigma^2) = -\frac{1}{2}\left[(1+\nu)\ln\sigma^2 + \frac{(y-\mu)^2}{\sigma^2} + \frac{\nu v}{\sigma^2}\right] + C ,$$

where C represents terms that do not depend on the adjustable parameters μ or σ^2. [6]

2019 Exam Question 3(a)

3(a) [6 marks] The random variables y and v are independent, and therefore the likelihood function is the product of their pdfs:

$$L(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} \frac{\left(\frac{\nu}{2\sigma^2}\right)^{\nu/2}}{\Gamma(\nu/2)} v^{\nu/2-1} e^{-\nu v/2\sigma^2} \ .$$

The log-likelihood is therefore

$$\begin{aligned} \ln L(\mu, \sigma^2) &= -\frac{1}{2} \ln \sigma^2 - \frac{1}{2} \frac{(y-\mu)^2}{\sigma^2} + \alpha \ln \beta - \beta v + C \\ &= -\frac{1}{2} \left[(1+\nu) \ln \sigma^2 + \frac{(y-\mu)^2}{\sigma^2} + \frac{\nu v}{\sigma^2} \right] + C' \ , \end{aligned}$$

where C and C' represent terms that do not depend on μ or σ^2.

2019 Exam Question 3(b)

(b) Show that the maximum-likelihood estimators $\hat{\mu}$ and $\widehat{\sigma^2}$ and the profiled estimator $\widehat{\widehat{\sigma^2}}(\mu)$ are

$$\begin{aligned}
\hat{\mu} &= y \, , \\
\widehat{\sigma^2} &= \frac{\nu v}{1+\nu} \, , \\
\widehat{\widehat{\sigma^2}}(\mu) &= \frac{\nu v + (y-\mu)^2}{1+\nu} \, .
\end{aligned}$$

[10]

3(b) [10 marks] Setting the derivative of $\ln L$ with respect to σ^2 to zero gives

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{1}{2}\left[\frac{1+\nu}{\sigma^2} - \frac{(y-\mu)^2}{(\sigma^2)^2} - \frac{\nu v}{(\sigma^2)^2}\right] = 0 \, . \tag{1}$$

2019 Exam Question 3(b)

Solving for σ^2 as a function of μ gives the profiled value

$$\widehat{\widehat{\sigma^2}}(\mu) = \frac{\nu v + (y-\mu)^2}{1+\nu} \ .$$

Setting the derivative of $\ln L$ with respect to μ to zero gives

$$\frac{\partial \ln L}{\partial \mu} = \frac{y-\mu}{\sigma^2} = 0 \ . \qquad (2)$$

Solving Eqs. (1) and (2) simultaneously for μ and σ^2 give the maximum-likelihood estimators

$$\begin{aligned} \hat{\mu} &= y \ , \\ \widehat{\sigma^2} &= \frac{\nu v}{1+\nu} \ . \end{aligned}$$

2019 Exam Question 3(c)

(c) Suppose we use the statistic

$$t_\mu = -2 \ln \frac{L(\mu, \widehat{\widehat{\sigma^2}}(\mu))}{L(\hat{\mu}, \widehat{\sigma^2})}$$

to test hypothesized values of μ. Using the ingredients found above, show that t_μ is

$$t_\mu = (1+\nu) \ln \left[1 + \frac{1}{\nu} \frac{(y-\mu)^2}{v} \right] .$$

[8]

The parametric form of $f_v(v)$ was chosen such that the expectation value and variance of v are $E[v] = \sigma^2$ and $V[v] = 2\sigma^4/\nu$, i.e., v is an estimate of the variance σ^2. Let $s = \sqrt{v}$ be the corresponding estimator for σ.

3(c) [8 marks] Using the ingredients found in (a) and (b) we can find the profile log-likelihood,

$$\ln L(\mu, \widehat{\widehat{\sigma^2}}) = -\frac{1}{2} \left[(1+\nu) \ln \frac{\nu v + (y-\mu)^2}{1+\nu} + \frac{(y-\mu)^2 (1+\nu)}{\nu v + (y-\mu)^2} + \frac{\nu v (1+\nu)}{\nu v + (y-\mu)^2} \right] + C$$

2019 Exam Question 3(c)

and also the maximum of the log-likelihood,

$$\ln L(\hat{\mu}, \widehat{\sigma^2}) = -\frac{1}{2}(1+\nu)\left[1 + \ln\frac{v\nu}{1+\nu}\right] + C \ ,$$

where the constants C are the same in both expressions above. The statistic t_μ is therefore

$$\begin{aligned} t_\mu &= -2\ln\frac{L(\mu, \widehat{\widehat{\sigma^2}})}{L(\hat{\mu}, \widehat{\sigma^2})} = (1+\nu)\ln\frac{\nu v + (y-\mu)^2}{1+\nu}\frac{1+\nu}{\nu v} + \frac{(1+\nu)[(v\nu + (y-\mu)^2]}{v\nu + (y-\mu)^2} - (1+\nu) \\ &= (1+\nu)\ln\left[1 + \frac{1}{\nu}\frac{(y-\mu)^2}{v}\right] \ . \end{aligned}$$

2019 Exam Question 3(d)

(d) Assuming that one approximates $E[v] \approx (E[s])^2$, show using error propagation that the ratio of the standard deviation of s to its mean is

$$\frac{\sigma_s}{E[s]} \approx \frac{1}{\sqrt{2\nu}} \, .$$

[6]

3(d) [6 marks] The estimate s of σ is $s = v^{1/2}$. Using error propagation to find the standard deviation of s gives

$$\sigma_s = \left| \frac{\partial s}{\partial v} \right|_{E[v]} \sigma_v = \left. \frac{1}{2} v^{-1/2} \right|_{E[v]} \sigma_v = \frac{1}{2} \frac{\sigma_v}{\sqrt{E[v]}} \, .$$

Using $E[v] = \sigma^2$, $V[v] = 2\sigma^4/\nu$ and $E[s] \approx \sqrt{E[v]}$ (all given) we find

$$\frac{\sigma_s}{E[s]} \approx \frac{\sigma_s}{\sqrt{E[v]}} = \frac{1}{2} \frac{\sigma_v}{E[v]} = \frac{1}{2} \sqrt{\frac{2}{\nu}} \sigma^2 \frac{1}{\sigma^2} = \frac{1}{\sqrt{2\nu}} \, .$$

2019 Exam Question 3(e)

(e) Starting from the pdf $f_v(v)$ given above, find the pdf for s in terms of the parameters ν and σ^2. [5]

3(e) [5 marks] The pdf of s is

$$g(s) = \left| \frac{dv}{ds} \right| f(v(s)) \, ,$$

where the pdf for v is

$$f(v) = \frac{\left(\frac{\nu}{2\sigma^2} \right)^{\nu/2}}{\Gamma(\nu/2)} v^{\nu/2-1} e^{-\nu v/2\sigma^2} \, .$$

Using $v = s^2$ we have $dv/ds = 2s$ and therefore

$$g(s) = 2s \frac{\left(\frac{\nu}{2\sigma^2} \right)^{\nu/2}}{\Gamma(\nu/2)} s^{2(\nu/2-1)} e^{-\nu s^2/2\sigma^2} = 2 \frac{\left(\frac{\nu}{2\sigma^2} \right)^{\nu/2}}{\Gamma(\nu/2)} s^{\nu-1} e^{-\nu s^2/2\sigma^2} \, .$$

2019 Exam Question 3(f)

(f) Show that in the limit where ν is very large, the statistic t_μ becomes

$$t_\mu \approx \frac{(y-\mu)^2}{\sigma^2} \ .$$

[5]

3(f) [5 marks] Having $\nu \to \infty$ means $\sigma_v^2 = 2\sigma^4/\nu \to 0$, i.e., the estimate v is always equal to σ^2. Expanding the logarithm using $\ln(1+\varepsilon) \approx \varepsilon$ for small ϵ and replacing v by σ^2 gives

$$t_\mu = (1+\nu) \ln \left[1 + \frac{1}{\nu} \frac{(y-\mu)^2}{v} \right] \to \frac{(y-\mu)^2}{\sigma^2} \ .$$