Modern Methods of Statistical Data Analysis

From parameter estimation to deep learning — A guided tour of probability

Lecture 7

-

Confidence Intervals

P.-D. Dr. Roger Wolf
roger.wolf@kit.edu

Dr. Pablo Goldenzweig
pablo.goldenzweig@kit.edu

Dr. Jan Kieseler
jan.kieseler@kit.edu

Dr. Slavomira Stefkova
slavomira.stefkova@kit.edu

Today

- Classical confidence intervals
  - Exact method
  - Examples:
    - Gaussian distributed estimator
    - Poisson distributed estimator
  - Likelihood and LS confidence intervals
  - Multi-dimensional confidence regions

Evaluations: Lecture & Computerpraktikum.

Please take a few minutes to fill them out. Your feedback is greatly appreciated. We will take your comments into consideration in trying to improve the course.

Evaluation period: through 22 June (lecture) & 15 July (Computerpraktikum)

Answer Time: Quiz 6

Type I versus *Type II* errors

1. You have two hypotheses H_0 and H_1 and a test statistics t distributed according to $g(t|H_0)$ and $g(t|H_1)$, as shown in the figure below. You now choose a certain value t_{cut} to accept H_0 / reject H_0. Using the figure, explain the meaning of Type I and Type II errors.

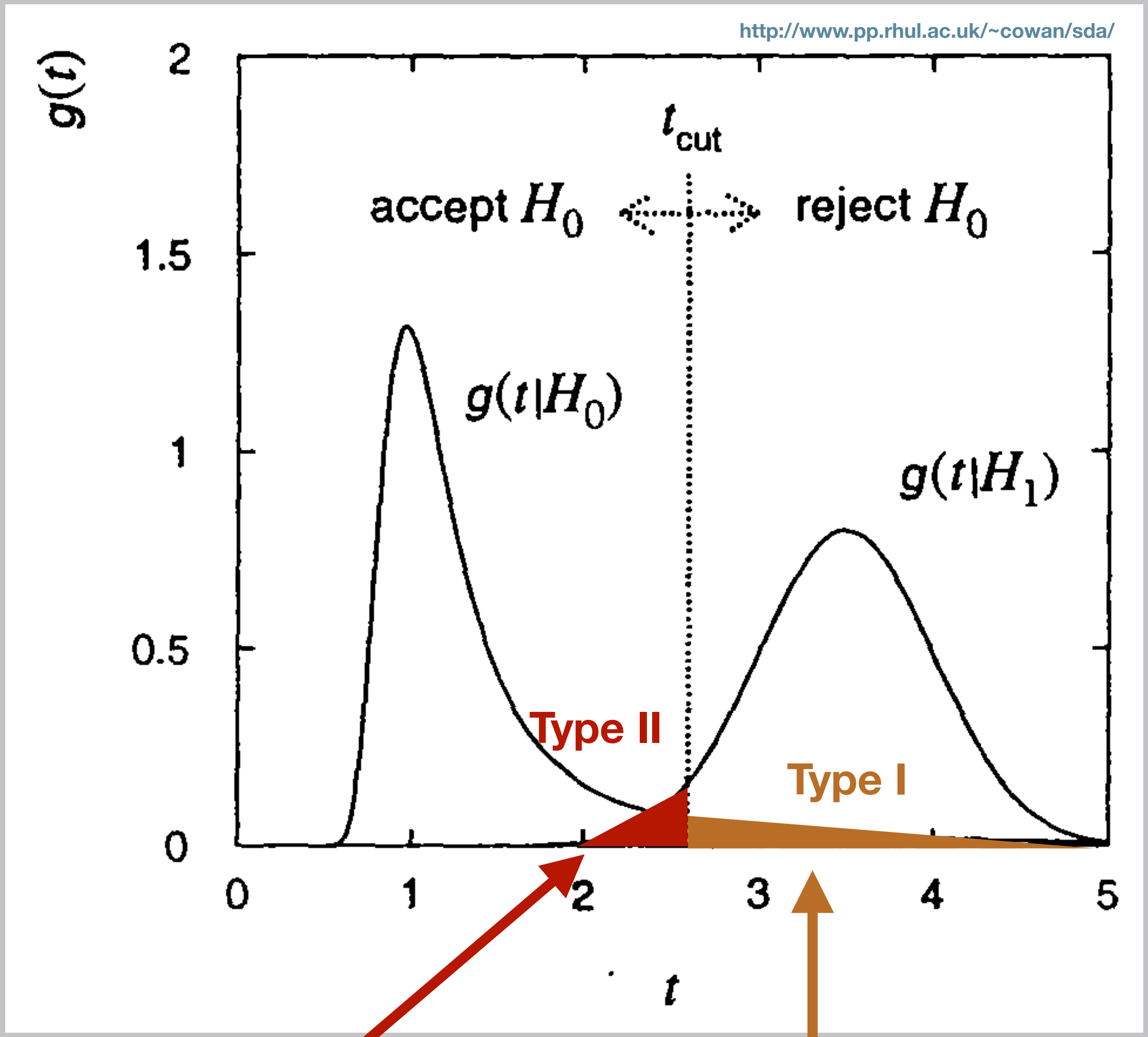


Table of error types		Null hypothesis (H_0) is	
		True	False
Decision About Null Hypothesis (H_0)	Fail to reject	Correct inference (True Negatives)	Type II error (False Negative)
	Reject	Type I error (False Positive)	Correct inference (True Positives)

Type I $\alpha = \int_{t_{\text{cut}}}^{\infty} g(t|H_0)\,dt.$

Type II $\beta = \int_{-\infty}^{t_{\text{cut}}} g(t|H_1)\,dt.$

Accepting H_0 although not true ($t < t_{\text{cut}}$)

The probability of α to reject H_0, even if H_0 is true ($t > t_{\text{cut}}$)

Fisher etc.

a) Write down the definition of a Fisher discriminant of n data points $x = (x_1, x_2, \ldots, x_n)$.

$$t(\boldsymbol{x}) = \sum_{i=1}^{n} a_i x_i = \boldsymbol{a}^T \boldsymbol{x}$$

b) When is it beneficial to construct a Fisher discriminant instead using the full likelihood ratio?

Computationally easier to construct, especially for high dimensional PDFs that need MC simulation

c) You carried out least square fits (LS) using two hypotheses and obtained χ_0^2 and χ_1^2. What is the equivalent of the likelihood ratio for binned data?

See Sec. 6.11 (Testing goodness-of-fit with ML)

$$-2 \ln \left(\frac{\mathscr{L}_1}{\mathscr{L}_0} \right) = \chi_1^2 - \chi_0^2 = \Delta\chi^2,$$ where for hypothesis i, $$\chi_i^2 = (\boldsymbol{x}_{\text{data}} - \boldsymbol{x}_i)\ \boldsymbol{C}_i^{-1}\ (\boldsymbol{x}_{\text{data}} - \boldsymbol{x}_i)$$

d)* Explain step-by-step how you would obtain $g(t|H_0)$ and $g(t|X_1)$ using MC techniques

Generate pseudo-experiments for x according to H_0 or H_1, then calculate $t(x)$ for each pseudo-experiment and produce histograms for both sets. The histograms will be proportional to $g(t\,|\,H_0)$ and $g(t\,|\,H_1)$.

Parameter estimation

Review

Suppose: n observations $(x_1, x_2, \ldots, x_n)$ and PDF $f(x;\theta)$

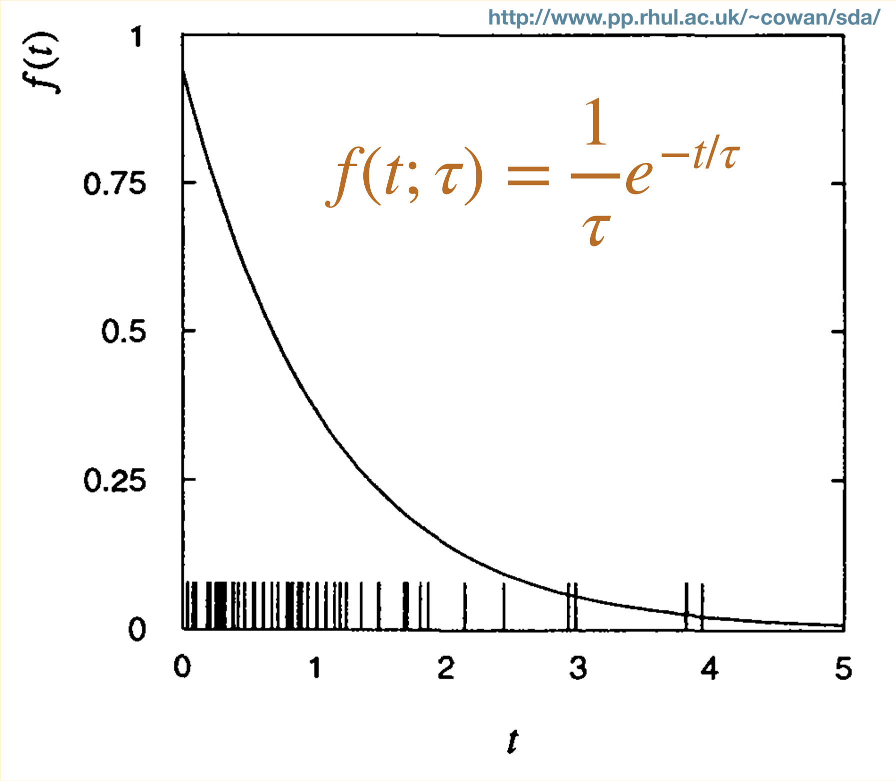
→ Construct $\hat{\theta}(x_1, x_2, \ldots, x_n)$ to estimate true value θ

→ $\hat{\theta}_{\text{obs}}$ = value of estimator actually observed

→ $\hat{\sigma}_{\hat{\theta}}$ = estimate of its standard deviation

Analytical method, MC, RCF bound (i)

- Discussed **3 methods** on estimating the variance of the found estimators
  - **Analytical method:**
    - Calculate the variance directly using the likelihood, e.g.



$$\log \mathscr{L}(\tau) = \sum_{i=1}^{n} \log f(t_i;\tau) = \sum_{i=1}^{n} \left(\log \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

$$V[\hat{\tau}] = E[\hat{\tau}^2] - (E[\hat{\tau}])^2$$

$$= \int \cdots \int \left(\frac{1}{n} \sum_{i=1}^{n} t_i \right)^2 \frac{1}{\tau} e^{-t_1/\tau} \cdots \frac{1}{\tau} e^{-t_n/\tau} \, dt_1 \ldots dt_n$$

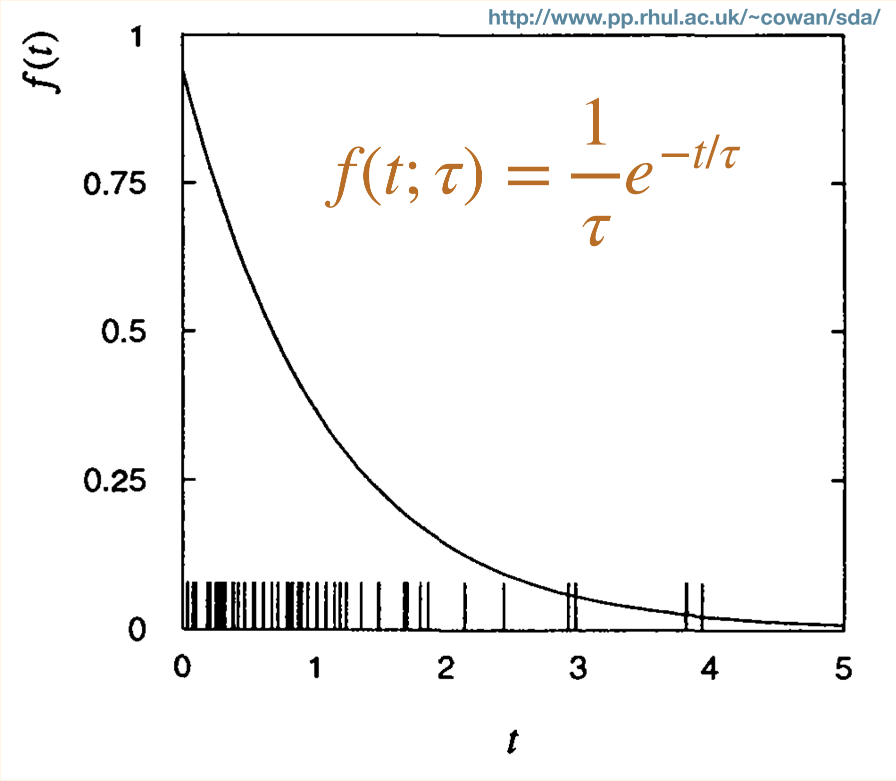
$$- \left(\int \cdots \int \left(\frac{1}{n} \sum_{i=1}^{n} t_i \right) \frac{1}{\tau} e^{-t_1/\tau} \cdots \frac{1}{\tau} e^{-t_n/\tau} \, dt_1 \ldots dt_n \right)^2 = \frac{\tau^2}{n}$$

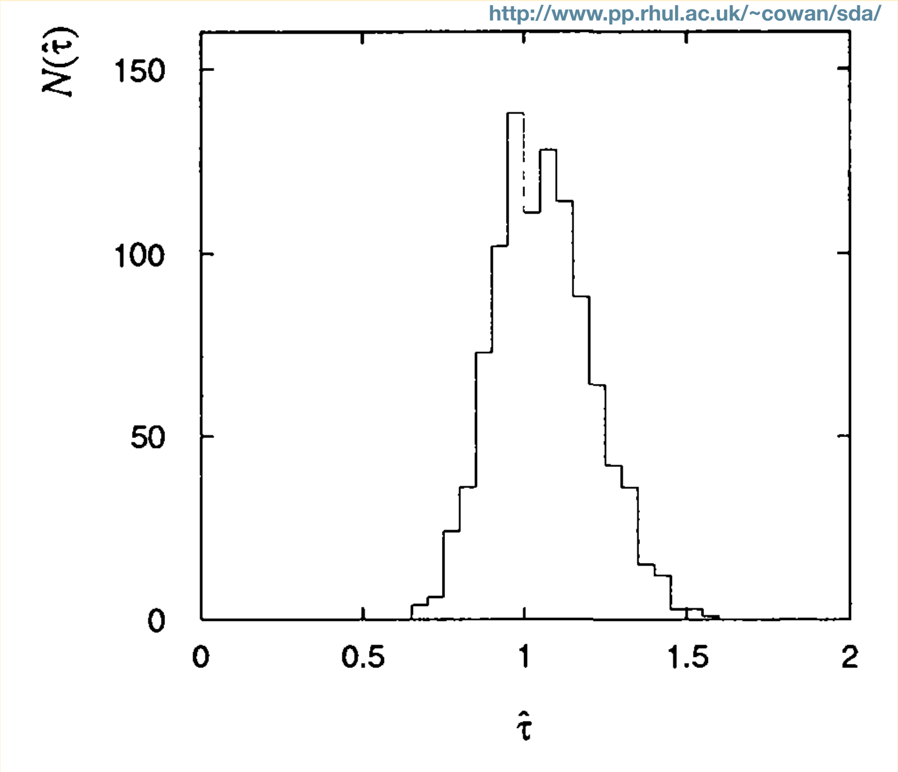
- **Often does not work.** If it works, can become **very complicated**: a priori trivial changes in the fit function (e.g. linear to quadratic PDF) result in you needing to recalculate complicated expressions

Analytical method, MC, RCF bound (ii)

Review

- Discussed **3 methods** on estimating the variance of the found estimators
  - **MC Method**
    - Simulate a large number of experiments, compute the ML or LS estimates



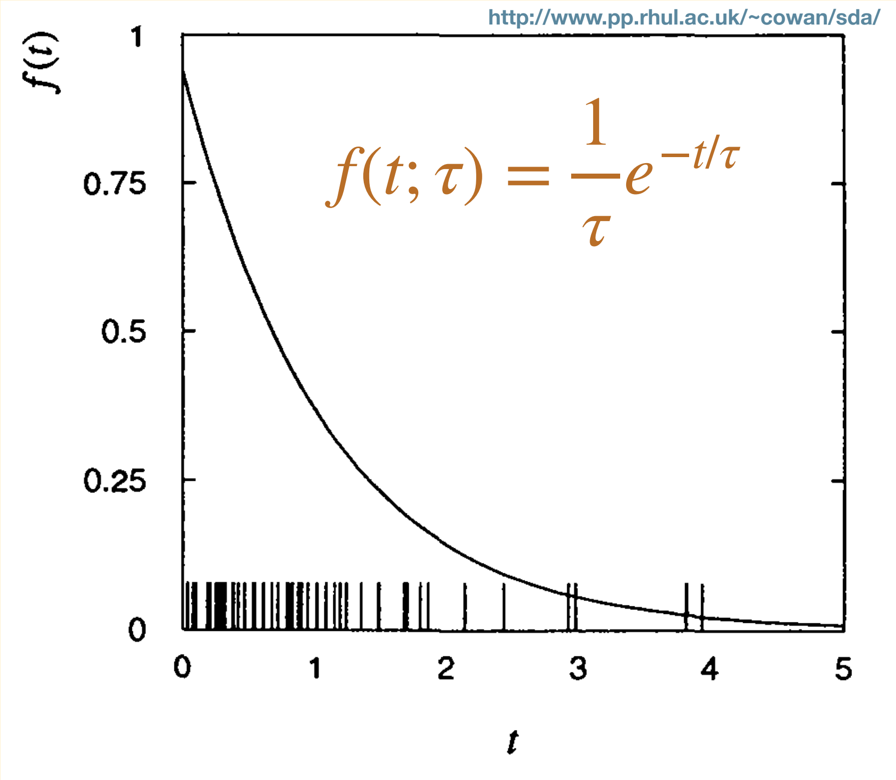


- **Simple**. Often a **good cross-check** for variances obtained via other methods (e.g. RCF/graphical) if validity is in question; can become too computationally expensive though.

Analytical method, MC, RCF bound (iii)

Review

- Discussed **3 methods** on estimating the variance of the found estimators
  - **RCF/graphical method**



Rao-Cramer-Frechet (RCF) inequality

$$V[\hat{\theta}] \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{E\left[-\frac{\partial^2 \log \mathscr{L}}{\partial \theta^2}\right]}$$

for bias free and efficient estimators

$$\left(\hat{V^{-1}}\right)_{ij} = \frac{\partial^2 \log \mathscr{L}}{\partial \theta_i \partial \theta_j}\bigg|_{\theta=\hat{\theta}}$$

$$\hat{\sigma}^2_{\hat{\theta}} = \left(-1 \Big/ \frac{\partial^2 \log \mathscr{L}}{\partial \theta^2}\right)\bigg|_{\theta=\hat{\theta}}$$

- Very nice **graphical** way to obtain the variance from the likelihood or LS curves

Analytical method, MC, RCF bound (iv)

- Discussed **3 methods** on estimating the variance of the found estimators
  - **RCF/graphical method**

Taylor expand around maximum

$$\log\mathscr{L}(\theta) = \log\mathscr{L}(\hat{\theta}) + \left[\frac{\partial \log\mathscr{L}}{\partial\theta}\right]_{\theta=\hat{\theta}}(\theta-\hat{\theta}) + \frac{1}{2!}\left[\frac{\partial^2 \log\mathscr{L}}{\partial\theta^2}\right]_{\theta=\hat{\theta}}(\theta-\hat{\theta})^2 + \dots$$

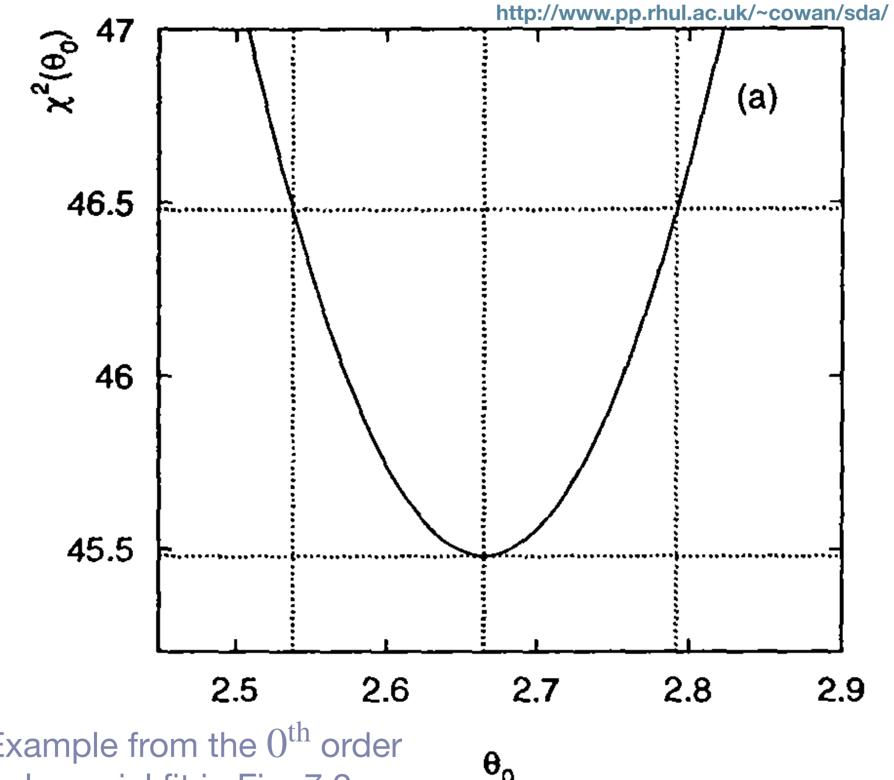
LS

$$\chi^2(\boldsymbol{\theta}) = \chi^2(\hat{\boldsymbol{\theta}}) + 1 = \chi^2_{\min} + 1$$

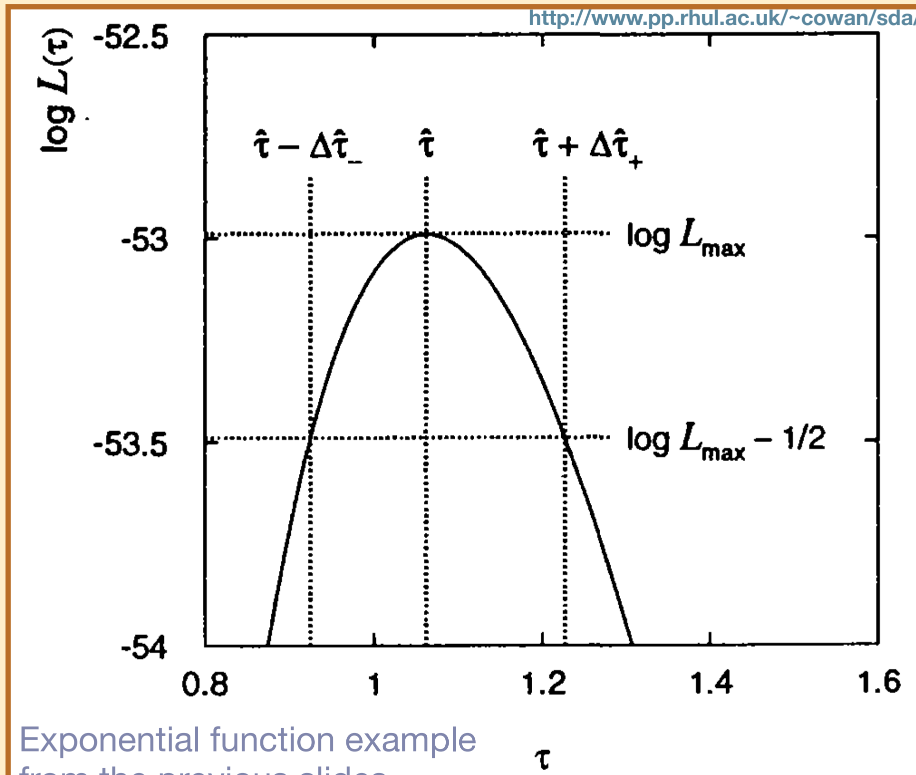
For the case of $\lambda(x;\boldsymbol{\theta})$ linear in the parameters $\boldsymbol{\theta}$

Likelihood

$$\log\mathscr{L}(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) = \log\mathscr{L}_{\max} - \frac{1}{2}$$



Example from the 0^{th} order polynomial fit in Fig. 7.2



Exponential function example from the previous slides

Analytical method, MC, RCF bound (v)

Review

Report $\hat{\theta}_{\rm obs} \pm \hat{\sigma}_{\hat{\theta}}$ as a proxy for $\theta \pm \sigma_\theta$ distributed as $g(\hat{\theta};\theta)$

sampling PDF

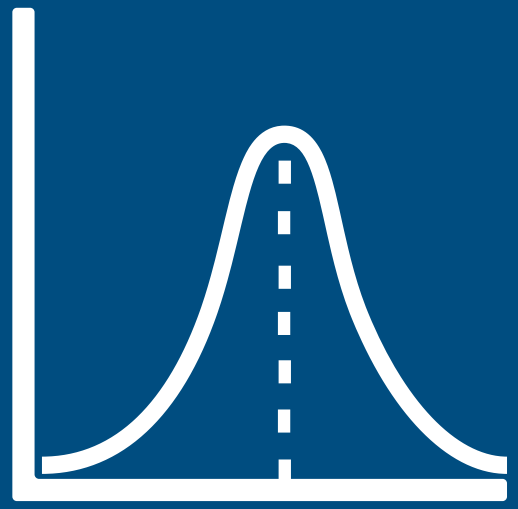
Interpret: repeated estimates all based on n observations of x would be distributed according to a PDF $g(\hat{\theta})$ centered around some true value θ and true standard deviation $\sigma_{\hat{\theta}}$, which are estimated to be $\hat{\theta}_{\rm obs}$ and $\hat{\sigma}_{\hat{\theta}}$

For most practical estimators: $g(\hat{\theta}) \to$ Gaussian in the large sample limit

But what if $g(\hat{\theta})$ is NOT Gaussian?

$\Rightarrow$ *Need to report confidence intervals (which can lead to asymmetric error bars)*

Confidence Intervals



Statistical errors, confidence intervals and limits

Up to now: when discussing 'error analysis' we focused on estimating the (co)variances of estimators. This is not always adequate and other ways of communicating the statistical uncertainty of measurements have to be found.

$$\hat{\theta}_{\mathrm{obs}} \pm \hat{\sigma}_{\hat{\theta}}$$

Classical confidence intervals (CI)

Alternative (& often =) method of reporting the statistical uncertainty of a measurement

- Suppose you have n observations of a random variable X, which can be used to evaluate an estimator for an unknown true parameter θ:

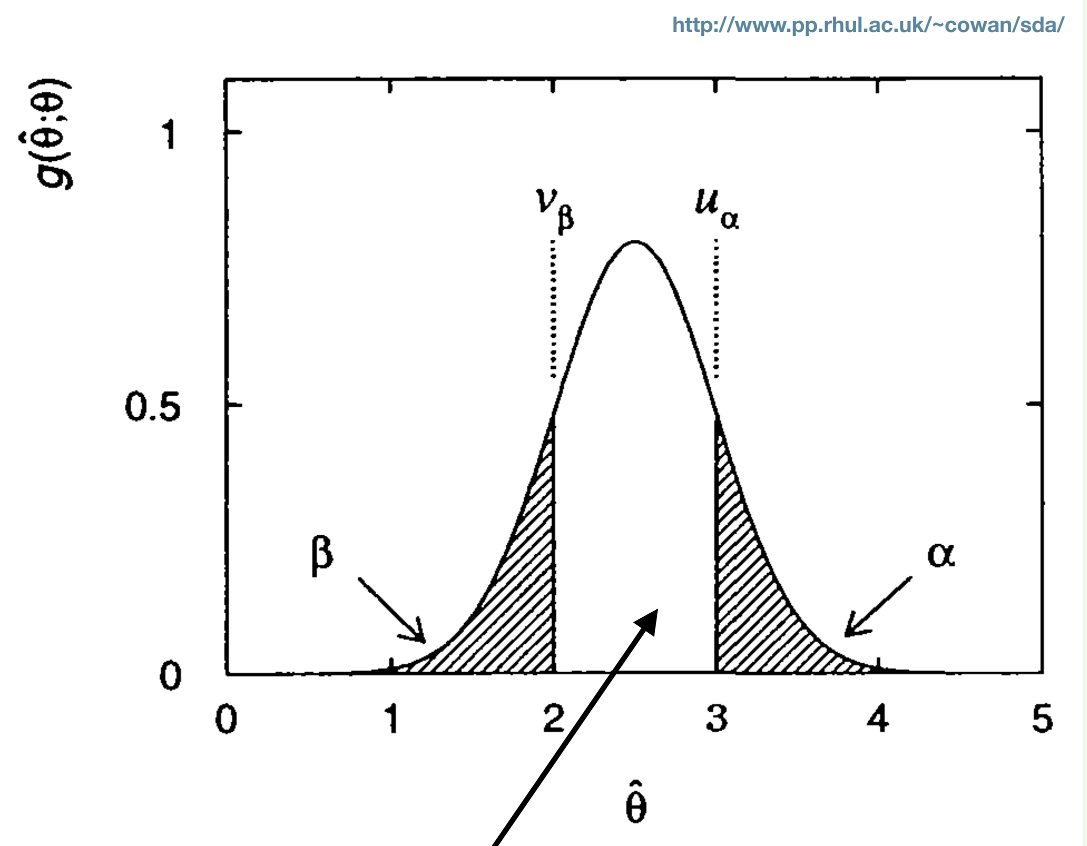
$$\hat{\theta}(x_1, \ldots, x_n) = \hat{\theta}_{\text{obs}}$$

↑ value obtained

- Furthermore, suppose we know the PDF of $\hat{\theta}$ denoted by $g(\hat{\theta}; \theta)$

  Real value of θ unknown, BUT for a given θ one knows what the PDF of $\hat{\theta}$ would be

- From $g(\hat{\theta}; \theta)$, can determine ν_β and u_α such that there are fixed probabilities β and α to observe $\hat{\theta} < \nu_\beta$ or $\hat{\theta} > u_\alpha$



$$P(v_\beta(\theta) \leq \hat{\theta} \leq u_\alpha(\theta)) = 1 - \alpha - \beta.$$

Shows the probability density for an estimator $\hat{\theta}$ for a particular value of the true parameter θ

u_α and u_β depend on the true value θ and are thus determined by

$$\beta = P(\hat{\theta} \leq v_\beta(\theta)) = \int_{-\infty}^{v_\beta(\theta)} g(\hat{\theta}; \theta) d\hat{\theta} = G(v_\beta(\theta); \theta),$$

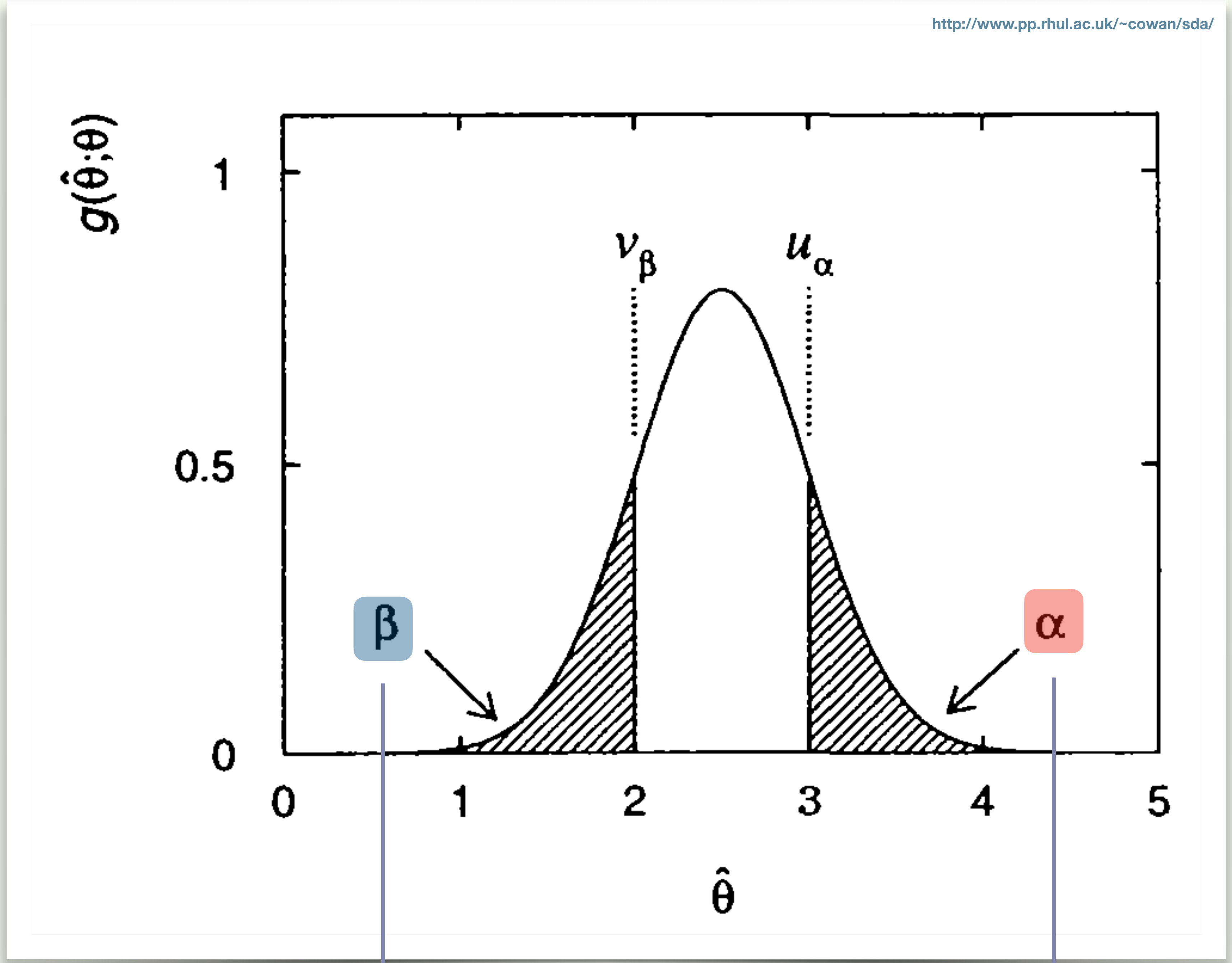
↑ CDF

$$\alpha = P(\hat{\theta} \geq u_\alpha(\theta)) = \int_{u_\alpha(\theta)}^{\infty} g(\hat{\theta}; \theta) d\hat{\theta} = 1 - G(u_\alpha(\theta); \theta),$$

... so α and β are the probabilities!

Next: lets build the CI step by step…

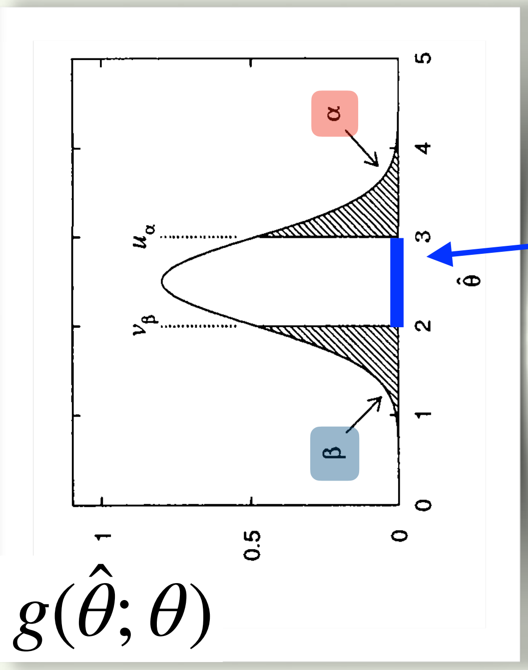
Confidence Belt (i)



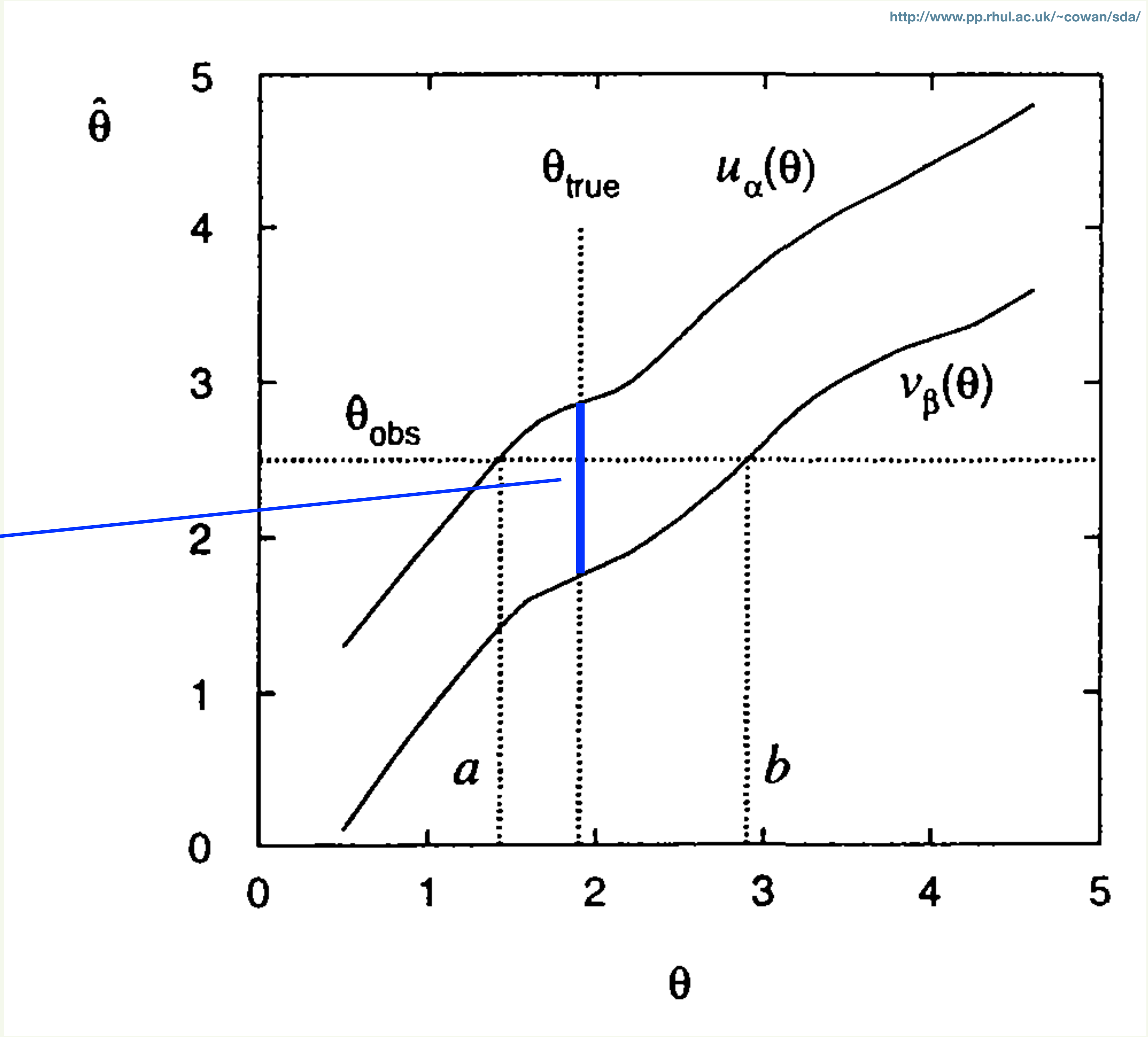
$$\beta = P(\hat{\theta} \leq v_\beta(\theta)) = \int_{-\infty}^{v_\beta(\theta)} g(\hat{\theta};\theta)d\hat{\theta} = G(v_\beta(\theta);\theta),$$

$$\alpha = P(\hat{\theta} \geq u_\alpha(\theta)) = \int_{u_\alpha(\theta)}^{\infty} g(\hat{\theta};\theta)d\hat{\theta} = 1 - G(u_\alpha(\theta);\theta),$$

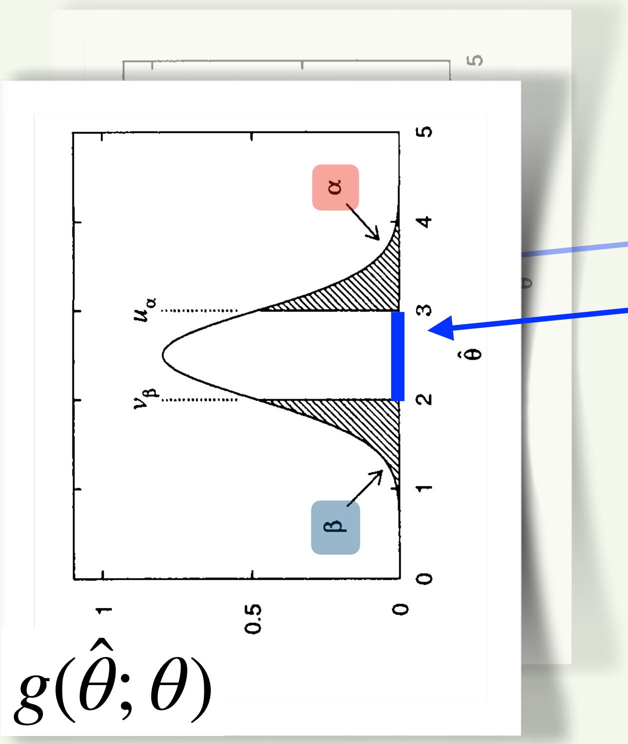
Confidence Belt (ii)

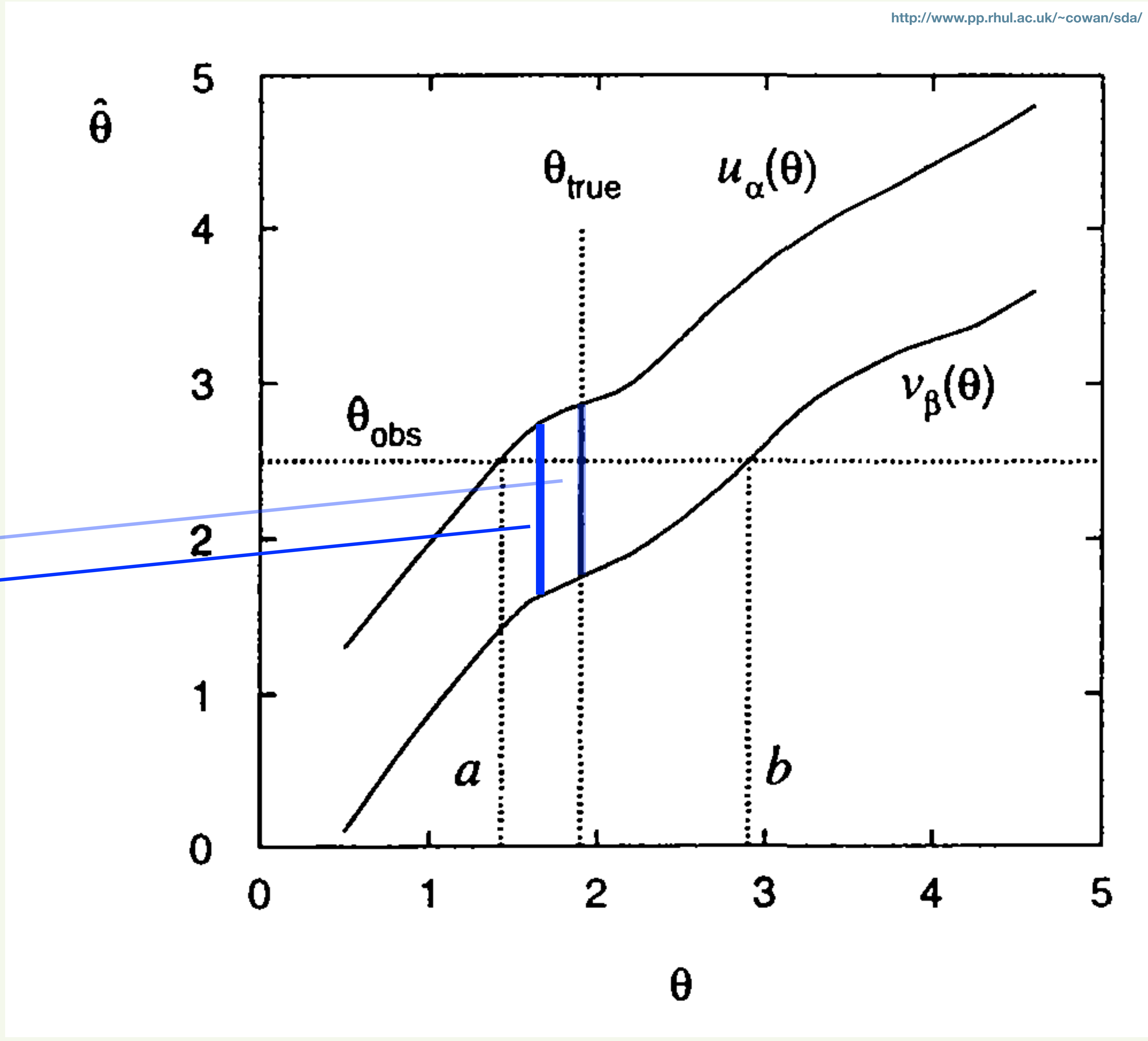


for a given value
of (true) θ



Confidence Belt (iii)





for *another* given value of (true) θ

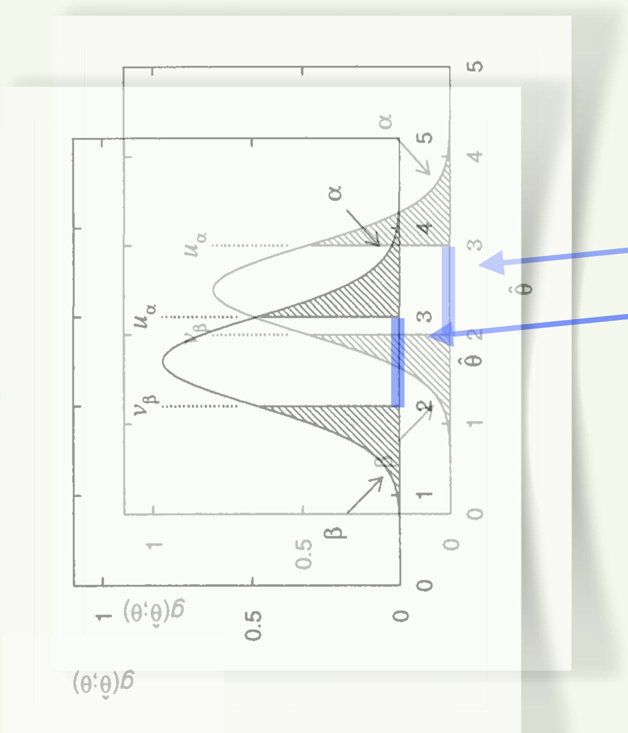
Confidence Belt (iv)

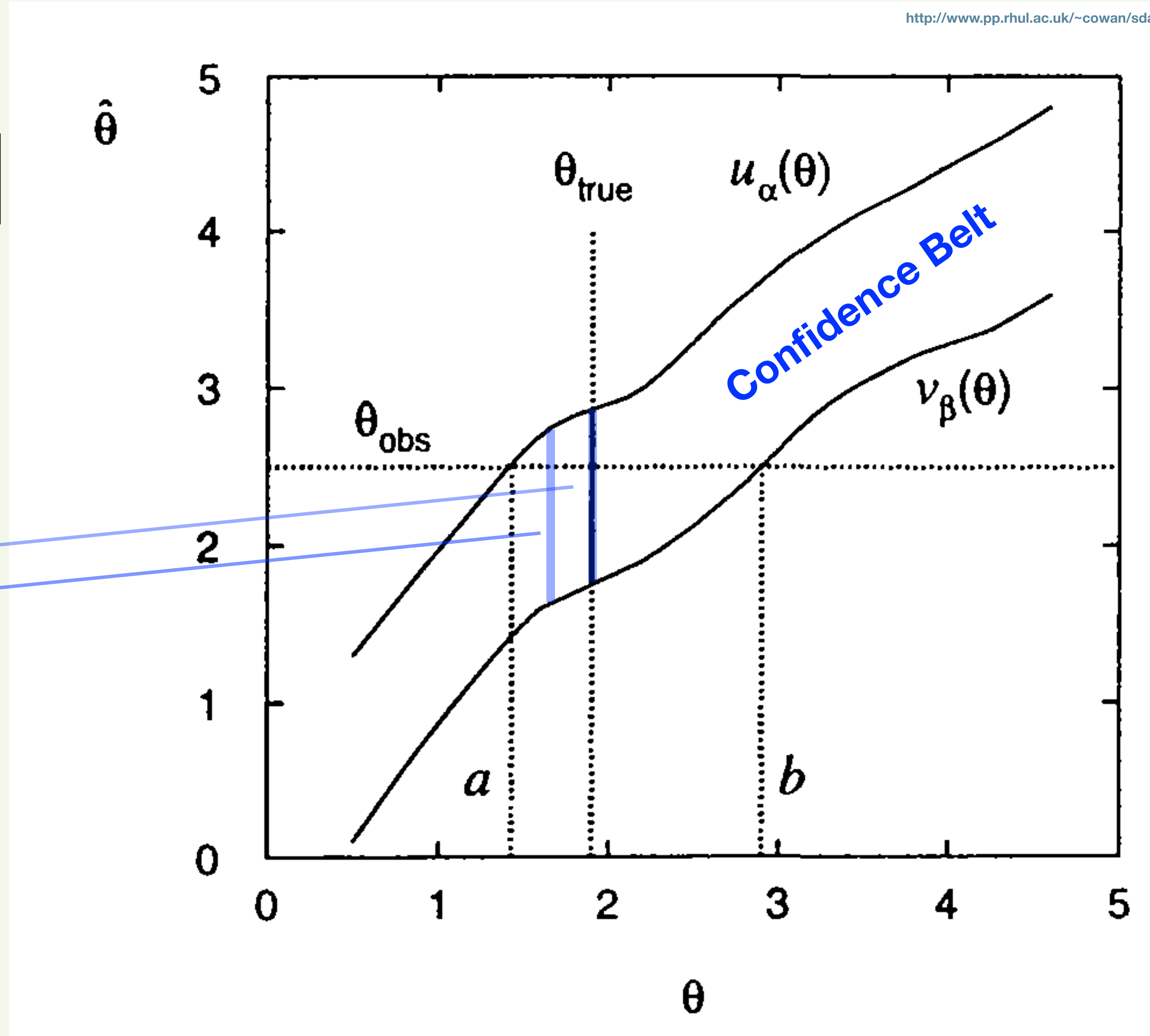
Region between the curves:

Confidence Belt
(Neyman Belt)

$$P(v_\beta(\theta) \leq \hat{\theta} \leq u_\alpha(\theta)) = 1 - \alpha - \beta.$$

The probability for the estimator $\hat{\theta}$ to be inside the belt, regardless of the value of θ



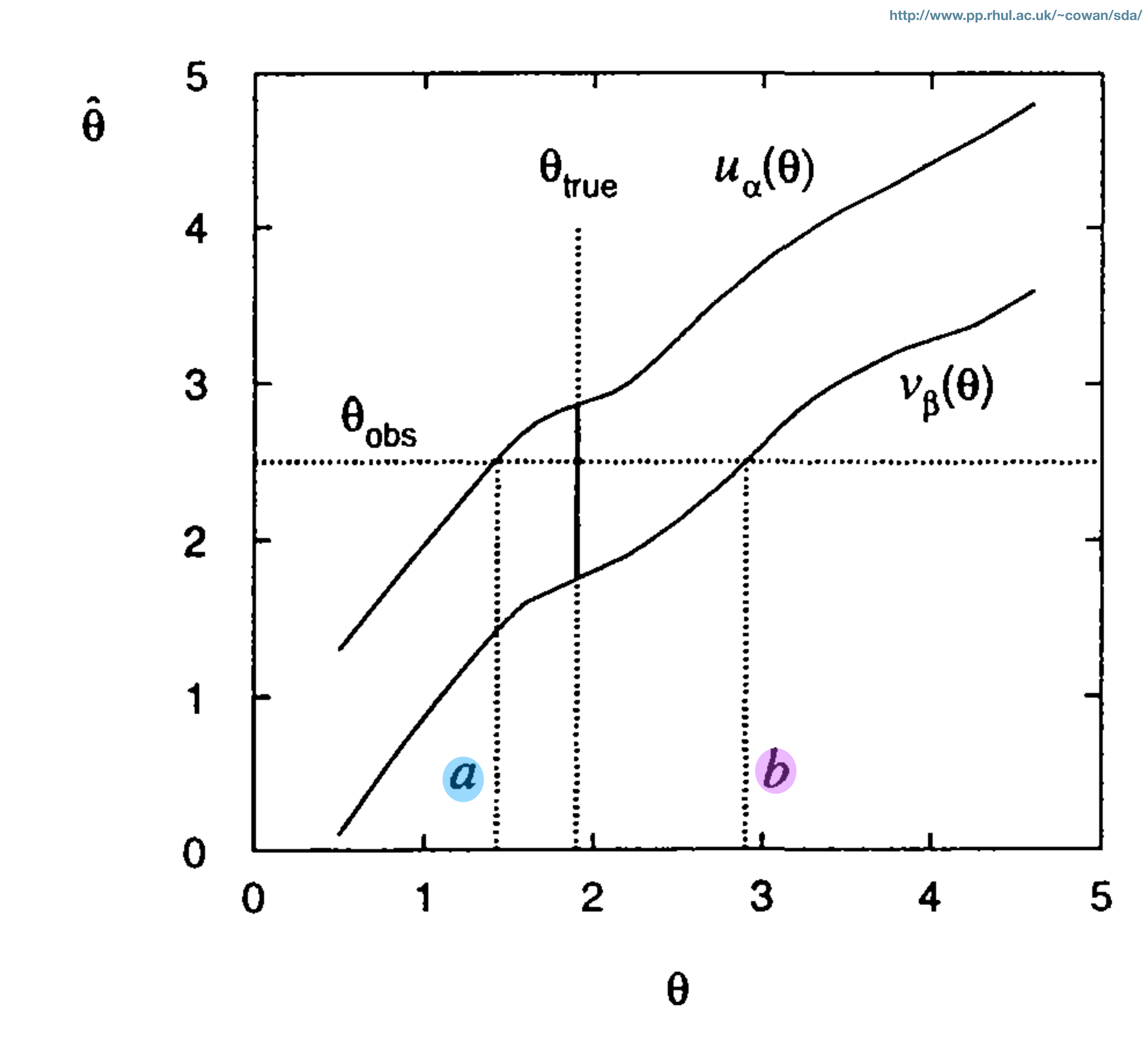


Confidence Interval (i)

If $u_\alpha(\theta)$ and $\nu_\beta(\theta)$ are monotonically increasing functions of θ, then one can determine the inverse functions

$$a(\hat{\theta}) \equiv u_\alpha^{-1}(\hat{\theta}),$$
$$b(\hat{\theta}) \equiv v_\beta^{-1}(\hat{\theta}).$$

(Should be the case if $\hat{\theta}$ is a good estimator for θ)



Confidence Interval (ii)

$$a(\hat{\theta}) \equiv u_\alpha^{-1}(\hat{\theta}),$$
$$b(\hat{\theta}) \equiv v_\beta^{-1}(\hat{\theta}).$$

This then implies:

$$\hat{\theta} \geq u_\alpha(\theta),$$
$$\hat{\theta} \leq v_\beta(\theta),$$

invert →

$$a(\hat{\theta}) \geq \theta,$$
$$b(\hat{\theta}) \leq \theta.$$

$$P(a(\hat{\theta}) \geq \theta) = \alpha,$$
$$P(b(\hat{\theta}) \leq \theta) = \beta,$$

$$P(v_\beta(\theta) \leq \hat{\theta} \leq u_\alpha(\theta)) = 1 - \alpha - \beta.$$

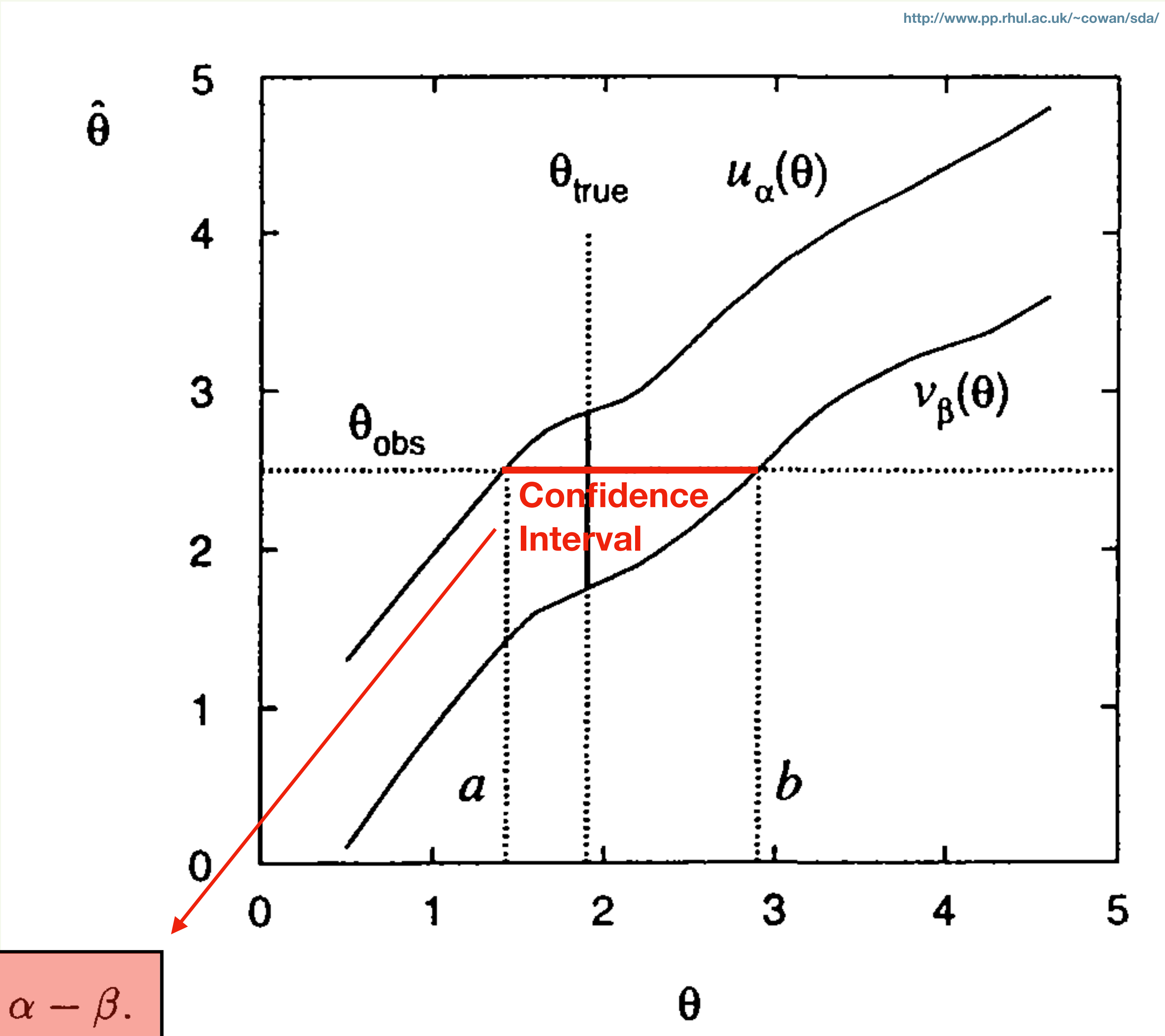
or

$$P(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta.$$

If the functions $a(\hat{\theta})$ and $b(\hat{\theta})$ are evaluated with the value of the estimator obtained in the experiment ($\hat{\theta}_{\text{obs}}$), then this determines 2 values $[a, b]$

Confidence Interval (iii)

Often chooses $\alpha = \beta = \frac{\gamma}{2}$ giving a so-called **central CI** with probability $= 1 - \gamma$



$[a, b]$: **Confidence Interval**, at a **confidence level** (or **coverage probability**) of $\mathbf{1 - \alpha - \beta}$

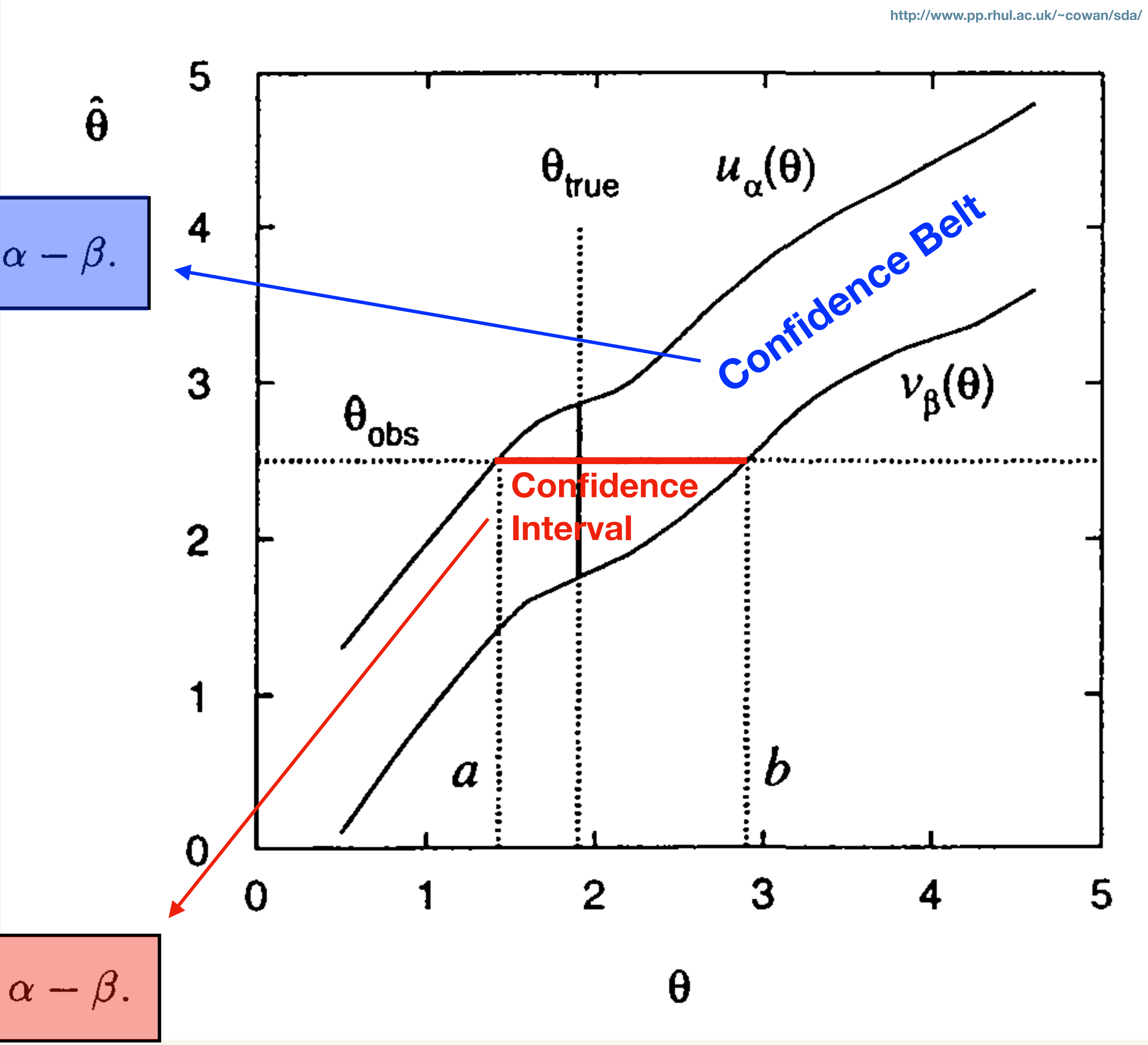
$$P(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta.$$

All together now

$$P(v_\beta(\theta) \leq \hat{\theta} \leq u_\alpha(\theta)) = 1 - \alpha - \beta.$$

Note where the ^s are in the 2 equations!

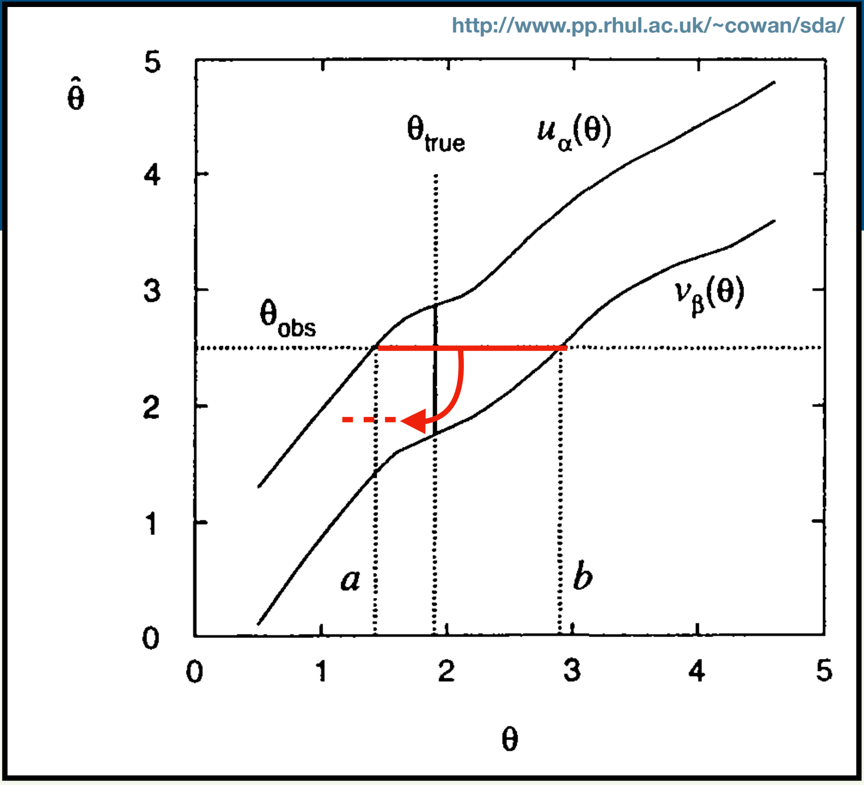
$[a, b]$: **Confidence Interval**, at a **confidence level** (or **coverage probability**) of $1 - \alpha - \beta$



$$P(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta.$$

Thus

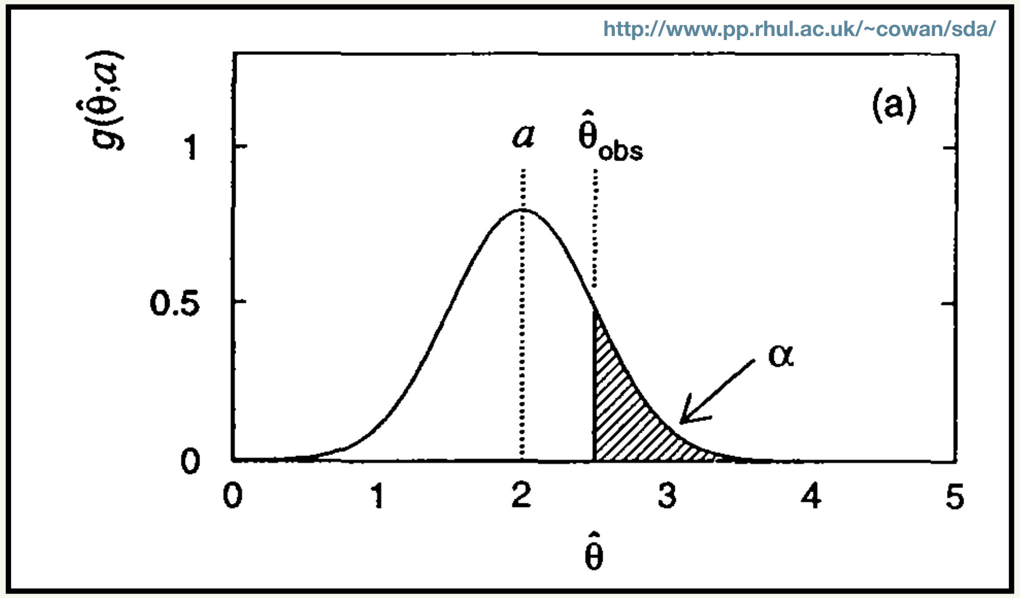
- The value a gives the *hypothetical value* of the true parameter θ for which a fraction α of repeated estimates $\hat{\theta}$ would be higher than the one actually obtained, $\hat{\theta}_{\rm obs}$
- Similarly the value b is the value of θ for which a fraction β of the estimates would be lower than $\hat{\theta}_{\rm obs}$

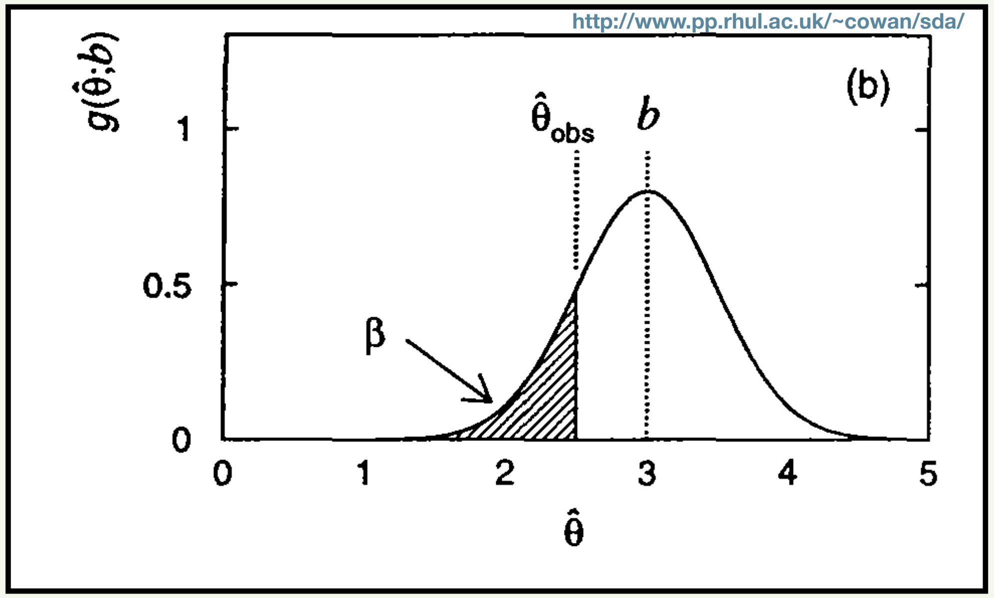


Taking $\hat{\theta}_{\rm obs} = u_\alpha(a) = \nu_\beta(b)$, the original equations become

$$\alpha = \int_{\hat{\theta}_{\rm obs}}^{\infty} g(\hat{\theta}; a)\, d\hat{\theta} = 1 - G(\hat{\theta}_{\rm obs}; a),$$

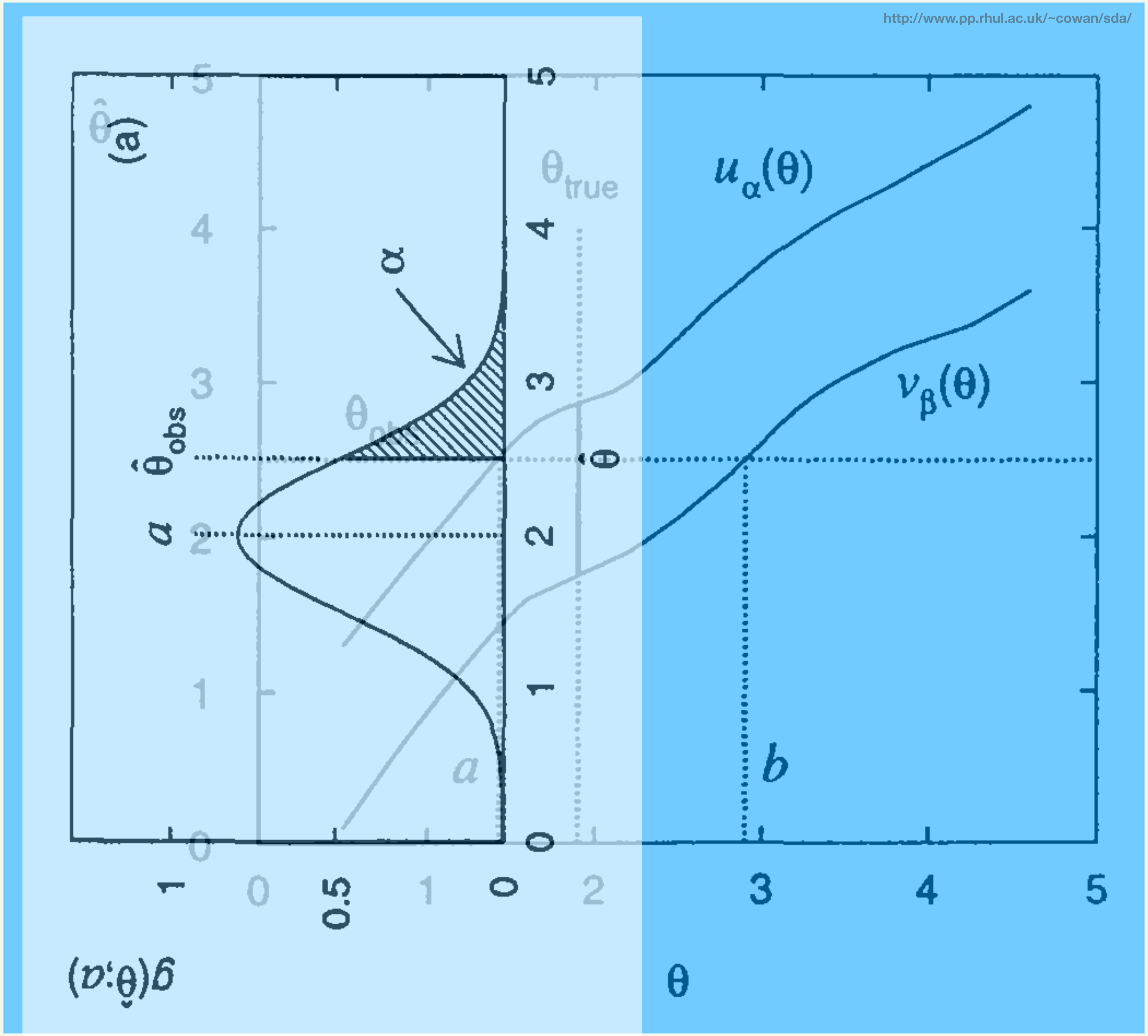
$$\beta = \int_{-\infty}^{\hat{\theta}_{\rm obs}} g(\hat{\theta}; b)\, d\hat{\theta} = G(\hat{\theta}_{\rm obs}; b).$$





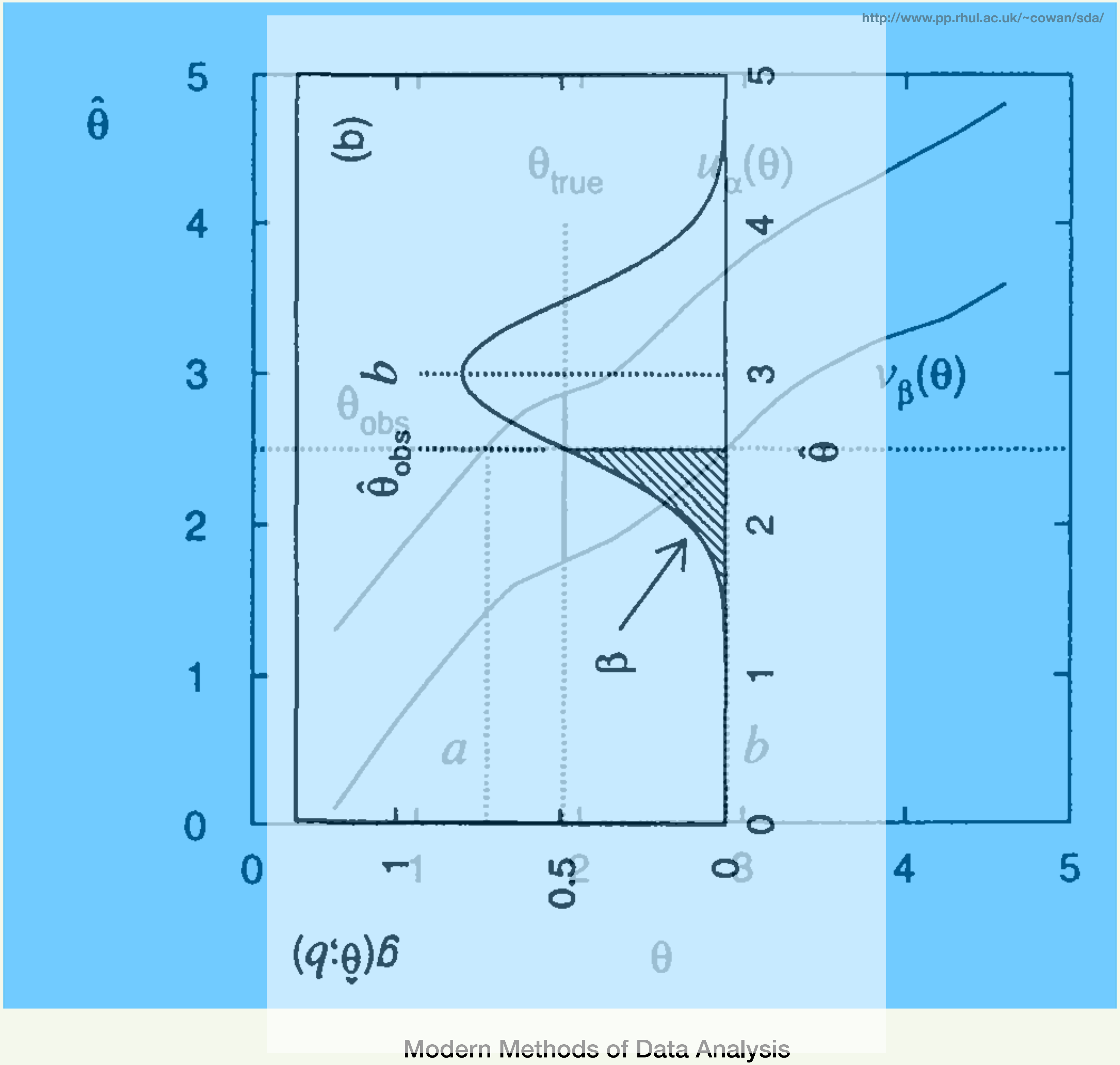
$$\alpha = \int_{\hat{\theta}_{obs}}^{\infty} g(\hat{\theta}; a)\, d\hat{\theta} = 1 - G(\hat{\theta}_{obs}; a),$$

Do you see α?



Do you see β?

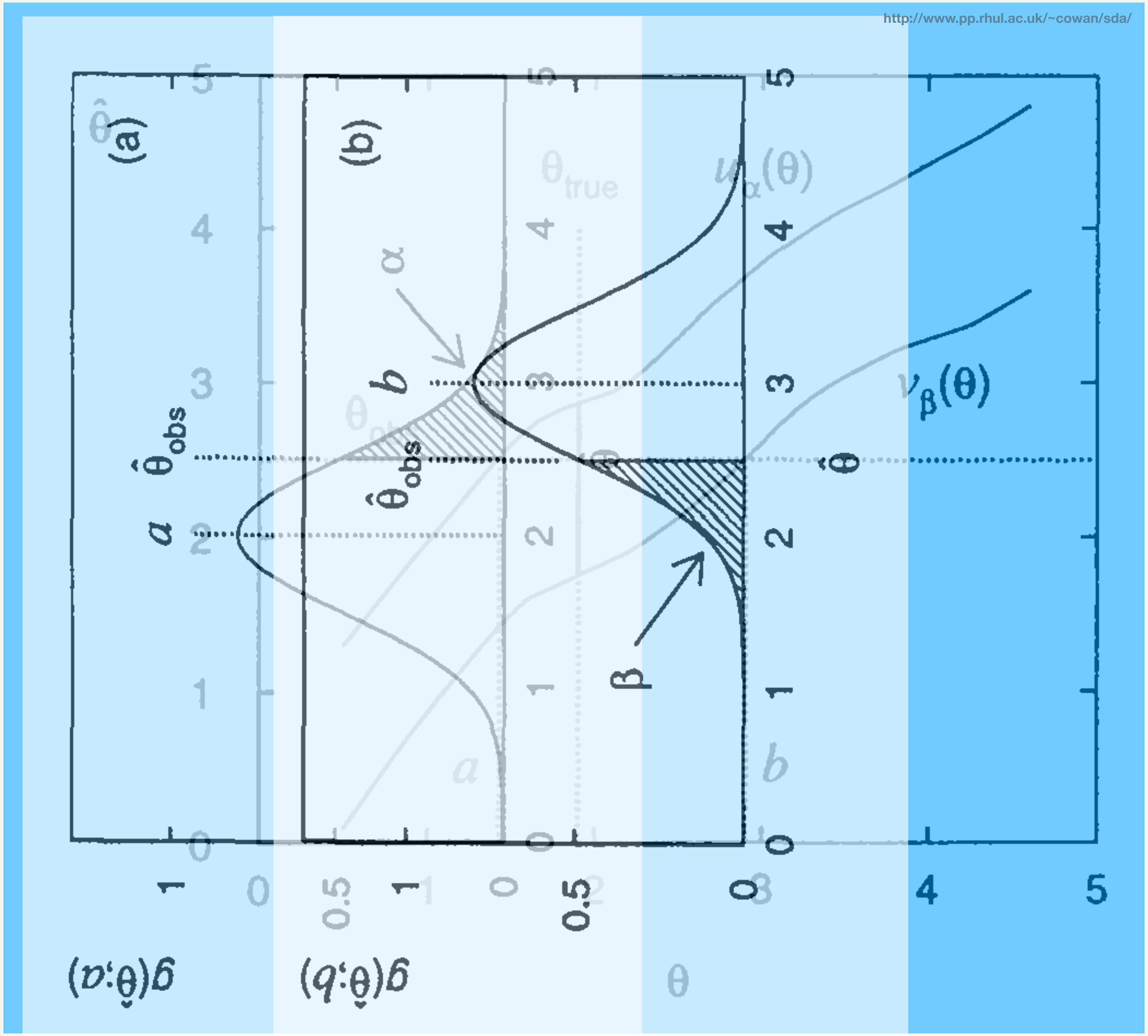
$$\beta = \int_{-\infty}^{\hat{\theta}_{\rm obs}} g(\hat{\theta}; b)\, d\hat{\theta} = G(\hat{\theta}_{\rm obs}; b).$$



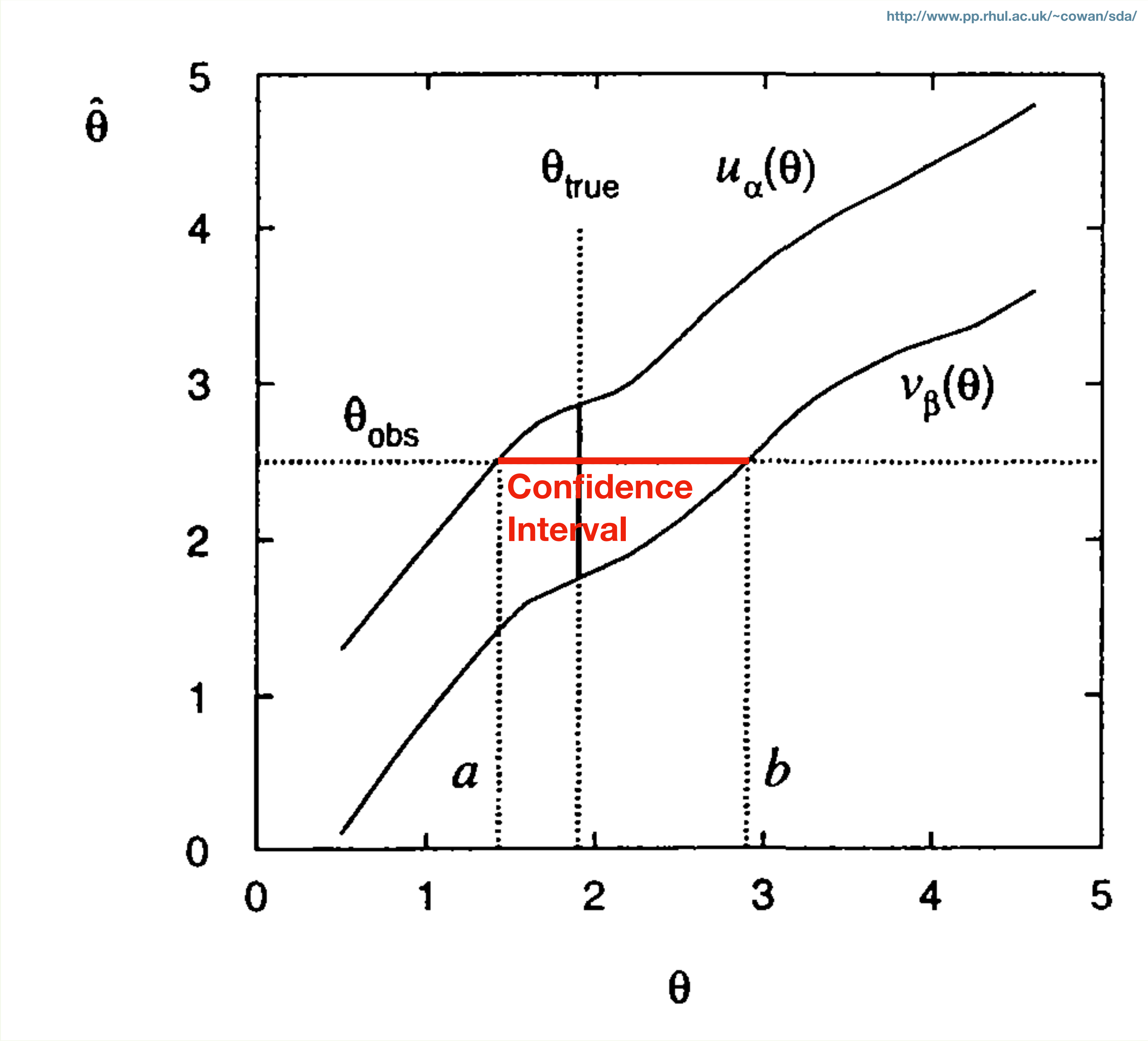
$$\alpha = \int_{\hat\theta_{obs}}^{\infty} g(\hat\theta; a)\, d\hat\theta = 1 - G(\hat\theta_{obs}; a),$$

Now both

$$\beta = \int_{-\infty}^{\hat\theta_{obs}} g(\hat\theta; b)\, d\hat\theta = G(\hat\theta_{obs}; b).$$



If the experiment were repeated many times, the interval $[a, b]$ would include the true value of the parameter θ in a fraction $1 - \alpha - \beta$ of the experiments



Error bars

- The confidence interval $[a, b]$ is often expressed by reporting the result of a measurement as →

often displayed as error bars

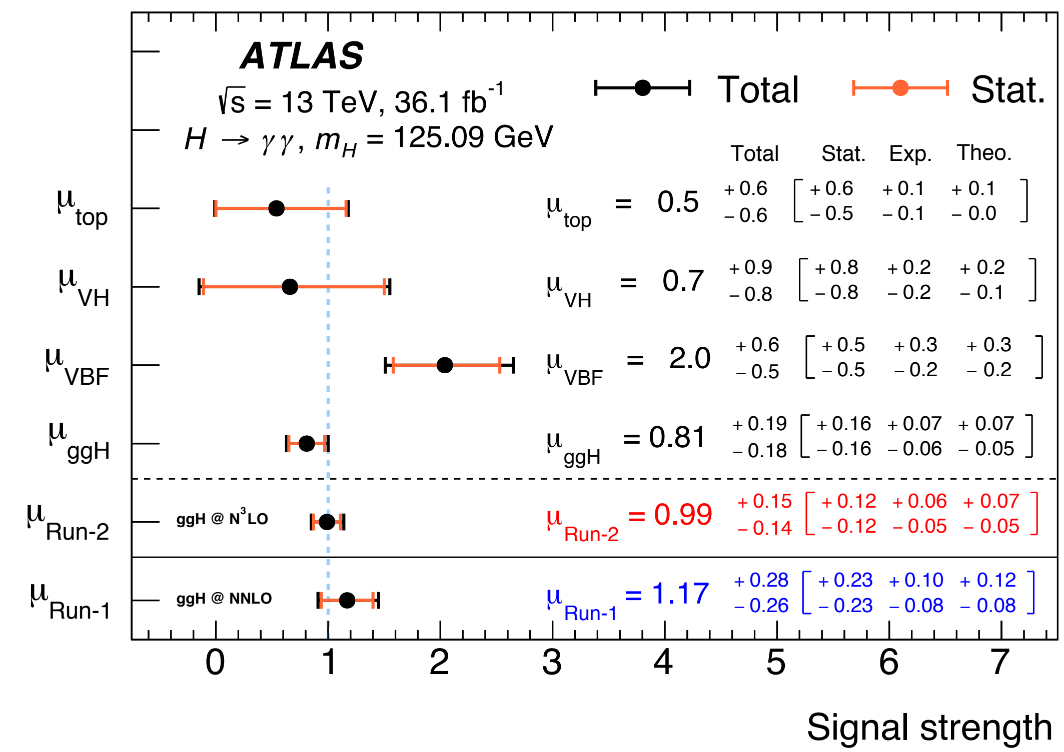
$$\hat{\theta}^{+d}_{-c}$$

$$d = b - \hat{\theta}$$

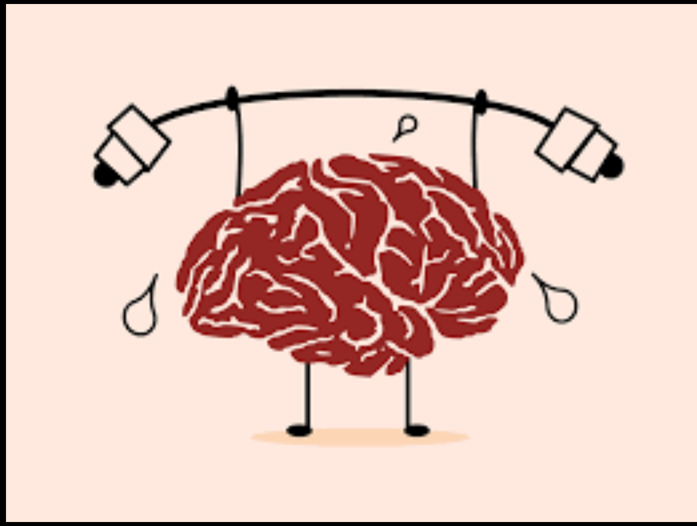
$$c = \hat{\theta} - a$$

estimated value

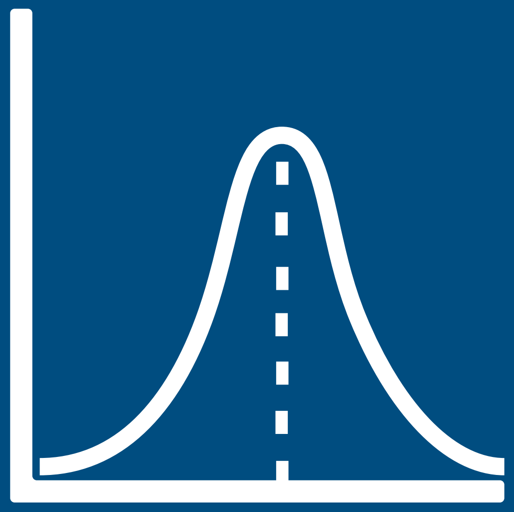
- In **many cases** the PDF $g(\hat{\theta}; \theta)$ is **approximatively** Gaussian, so that an interval of ± 1 one standard deviation around the measured value corresponds to a central confidence interval with $1 - \gamma = 0.683$.
  - *The 68.3% central confidence interval is usually adopted as the conventional definition for error bars even when the PDF of the estimator is not Gaussian*



Take 5



Gaussian Confidence Intervals



Let's apply what we've built up to the Gaussian limit

First recall how we compute probabilities with Normal RVs (Lecture 3, s14)

Review

- For a normal RV $X \sim \mathcal{N}(\mu, \sigma^2)$, its CDF has <u>no closed form</u>

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

- However, can solve for probabilities numerically using a function Φ

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

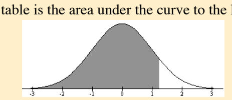
CDF of the Standard (unit) Normal, Z

Expectation: $E[Z] = \mu = 0$

Variance: $\mathrm{Var}[Z] = \sigma^2 = 1$

Standard Normal Table

Note: An entry in the table is the area under the curve to the left of z, $P(Z \leq z) = \Phi(z)$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

CI for Gaussian distributed estimators (i)

- **Simple and very important application:**
  - $\hat{\theta}$ is Gaussian with mean θ and standard deviation $\sigma_{\hat{\theta}}$ ⟵ Commonly occurring situation, since according to the CLT, any estimator that is a linear function of a sum of RVs becomes Gaussian in the large sample limit
  - Cumulative distribution of $\hat{\theta}$ is then

$$G(\hat{\theta}; \theta, \sigma_{\hat{\theta}}) = \int_{-\infty}^{\hat{\theta}} \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp\left(\frac{-(\hat{\theta}' - \theta)^2}{2\sigma_{\hat{\theta}}^2}\right) d\hat{\theta}' .$$

  - Suppose that the standard deviation is known and that the experiment resulted in an estimate $\hat{\theta}_{\rm obs}$. Then we can determine the confidence interval $[a, b]$ by solving

$$\alpha = 1 - G(\hat{\theta}_{\rm obs}; a, \sigma_{\hat{\theta}}) = 1 - \Phi\left(\frac{\hat{\theta}_{\rm obs} - a}{\sigma_{\hat{\theta}}}\right),$$

$$\beta = G(\hat{\theta}_{\rm obs}; b, \sigma_{\hat{\theta}}) = \Phi\left(\frac{\hat{\theta}_{\rm obs} - b}{\sigma_{\hat{\theta}}}\right),$$

⟵ **standard normal CDF** $\Phi = G(\hat{\mu}\ ;\ \mu = 0,\ \sigma = 1)$

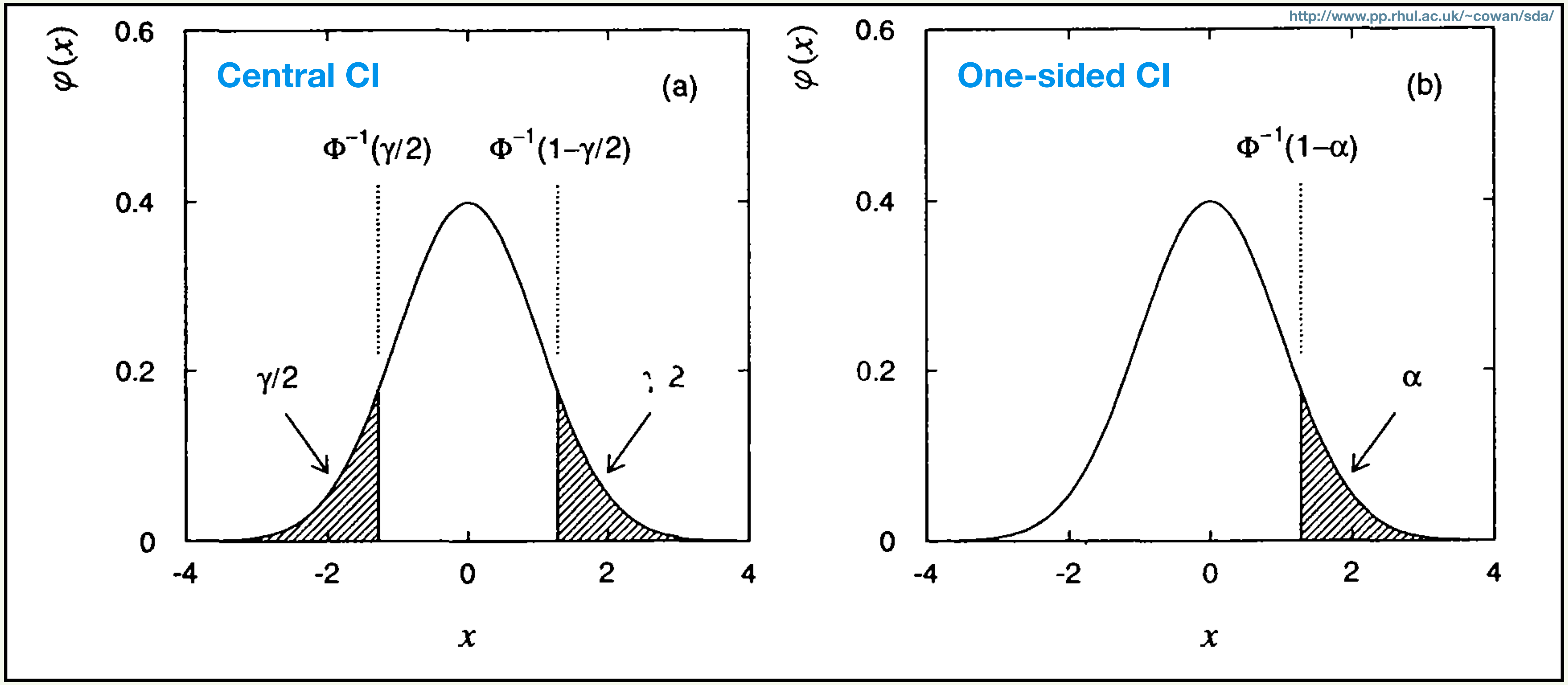
CI for Gaussian distributed estimators (ii)

- This results in

$$a = \hat{\theta}_{\text{obs}} - \sigma_{\hat{\theta}}\, \Phi^{-1}(1-\alpha),$$
$$b = \hat{\theta}_{\text{obs}} + \sigma_{\hat{\theta}}\, \Phi^{-1}(1-\beta).$$

inverse of standard normal CDF

i.e., the inverse function of Φ equals the quantile of the std. Gaussian



The relationship between the quantiles of the std. Gaussian distribution and the CI

CI for Gaussian distributed estimators (iii)

- Consider a central confidence interval with $\alpha = \beta = \gamma/2$
  - The confidence level $(1 - \gamma)$ is often chosen, such that $\Phi^{-1}(1 - \gamma/2)$ is a small integer (e.g., 1,2,3)
  - Similarly, one-sided intervals are often small integer values
    - Sometimes one also prefers to use a round value for $1 - \alpha$ or $1 - \gamma$

$\Phi^{-1}(1-\gamma/2)$	$1-\gamma$	$\Phi^{-1}(1-\alpha)$	$1-\alpha$
1	0.6827	1	0.8413
2	0.9544	2	0.9772
3	0.9973	3	0.9987

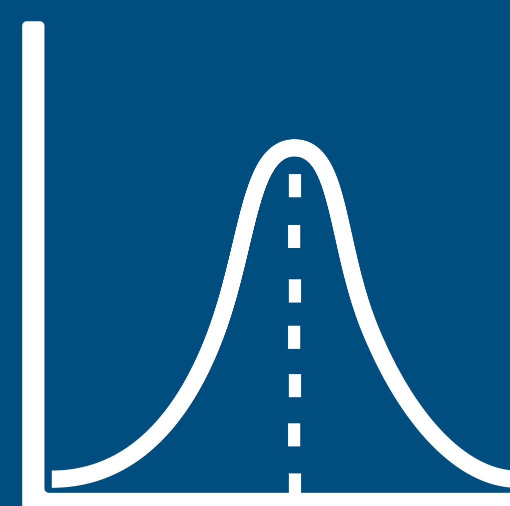
$1-\gamma$	$\Phi^{-1}(1-\gamma/2)$	$1-\alpha$	$\Phi^{-1}(1-\alpha)$
0.90	1.645	0.90	1.282
0.95	1.960	0.95	1.645
0.99	2.576	0.99	2.326

- For conventional 68.3% CI one has

$$[a, b] = [\hat{\theta}_{\text{obs}} - \sigma_{\hat{\theta}}, \hat{\theta}_{\text{obs}} + \sigma_{\hat{\theta}}].$$

- All of this is valid, if $\sigma_{\hat{\theta}}$ is known
  - Often not the case, *but in* ***large n limit*** *can use* $\sigma_{\hat{\theta}} \to \hat{\sigma}_{\hat{\theta}}$

Poisson Confidence Intervals



Let's look at cases where $\sigma_{\hat{\theta}} \rightarrow \hat{\sigma}_{\hat{\theta}}$ does not work

CI for Poisson distributed estimators (i)

- **The other common case:** Outcome of a measurement is a Poisson variable n $(n = 0,1,2, \ldots)$

  - Recall that the probability to observe n events is $f(n;\nu) = \frac{\nu^n}{n!}e^{-\nu}$
    - Expectation value: $E[n] = \nu$
    - Maximum Likelihood estimator: $\hat{\nu} = n$ → $\boxed{\hat{\nu}_{\text{obs}} = n_{\text{obs}}}$

Assume: single measurement and want to construct CI for ν

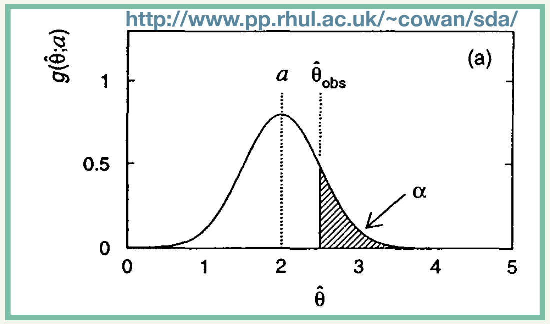
  - You will have some issues **directly** applying previous prescription: **only** integer values for $\hat{\nu}$ are possible, i.e. you cannot find $\hat{\nu}$ for arbitrary values of α, β such that:

$$u_\alpha \text{ with } P\left(\hat{\nu} \geq u_\alpha(\nu)\right) = \alpha$$

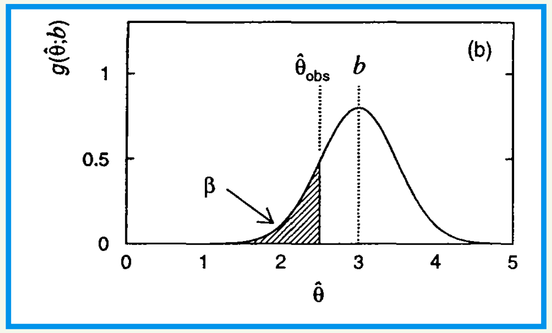
$$\nu_\beta \text{ with } P\left(\hat{\nu} \leq \nu_\beta(\nu)\right) = \beta$$

CI for Poisson distributed estimators (ii)

- However, the confidence interval $[a, b]$ can still be determined using:



$$\alpha = \int_{\hat{\theta}_{obs}}^{\infty} g(\hat{\theta}; a)\, d\hat{\theta} = 1 - G(\hat{\theta}_{obs}; a),$$



$$\beta = \int_{-\infty}^{\hat{\theta}_{obs}} g(\hat{\theta}; b)\, d\hat{\theta} = G(\hat{\theta}_{obs}; b).$$

$$\alpha = P(\hat{\nu} \geq \hat{\nu}_{obs}; a),$$
$$\beta = P(\hat{\nu} \leq \hat{\nu}_{obs}; b),$$

- For an estimate $\hat{\nu} = n_{obs}$ and given probabilities α and β, the following equations can be solved numerically for a and b:

$$\alpha = \sum_{n=n_{obs}}^{\infty} f(n; a) = 1 - \sum_{n=0}^{n_{obs}-1} f(n; a) = 1 - \sum_{n=0}^{n_{obs}-1} \frac{a^n}{n!} e^{-a},$$

$$\beta = \sum_{n=0}^{n_{obs}} f(n; b) = \sum_{n=0}^{n_{obs}} \frac{b^n}{n!} e^{-b}.$$

Next use the following relation between the Poisson and χ^2 distributions…

Connection between χ^2 and Poisson CDF

- There exists a useful relation between the Poisson and χ^2 distributions:

$$\begin{aligned}\sum_{n=0}^{n_{\mathrm{obs}}} \frac{\nu^n}{n!} e^{-\nu} &= \int_{2\nu}^{\infty} f_{\chi^2}(z; n_{\mathrm{d}} = 2(n_{\mathrm{obs}}+1))\, dz \\ &= 1 - F_{\chi^2}(2\nu; n_{\mathrm{d}} = 2(n_{\mathrm{obs}}+1)),\end{aligned}$$

  - Here f_{χ^2} is the χ^2 distribution with n_d degrees of freedom and F_{χ^2} is the corresponding cumulative distribution.

- Our two equations thus become

$$\begin{aligned} a &= \tfrac{1}{2} F^{-1}_{\chi^2}(\alpha; n_{\mathrm{d}} = 2n_{\mathrm{obs}}), \\ b &= \tfrac{1}{2} F^{-1}_{\chi^2}(1-\beta; n_{\mathrm{d}} = 2(n_{\mathrm{obs}}+1)).\end{aligned}$$

Lower and upper limits

Quantiles $F_{\chi^2}^{-1}$ of the χ^2 can be obtained from standard tables

- Example values for Poisson lower and upper limits for n_{obs} observed events:

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

n_{obs}	lower limit a			upper limit b		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0	–	–	–	2.30	3.00	4.61
1	0.105	0.051	0.010	3.89	4.74	6.64
2	0.532	0.355	0.149	5.32	6.30	8.41
3	1.10	0.818	0.436	6.68	7.75	10.04
4	1.74	1.37	0.823	7.99	9.15	11.60
5	2.43	1.97	1.28	9.27	10.51	13.11
6	3.15	2.61	1.79	10.53	11.84	14.57
7	3.89	3.29	2.33	11.77	13.15	16.00
8	4.66	3.98	2.91	12.99	14.43	17.40
9	5.43	4.70	3.51	14.21	15.71	18.78
10	6.22	5.43	4.13	15.41	16.96	20.14

Now look at the case where you have no observed events, but still want to set an upper limit

Poisson upper limits for $n_{obs} = 0$

- Very important special case:

$$\beta = \sum_{n=0}^{0} \frac{b^n e^{-b}}{n!} = e^{-b},$$

...or $b = -\log \beta$

& one is interested in establishing an upper limit b

- For the upper limit of a confidence level of $1 - \beta = 95\,\%$ one has $b = -\log(0.05) \approxeq 3$
- Thus if the number of occurrences of some rare event is treated as a Poisson variable with mean ν, and one looks for events of this type and finds none, then the 95% upper limit on the mean is 3.
- *That is, if the mean were in fact $\nu = 3$, the probability of observing zero would be 5%*

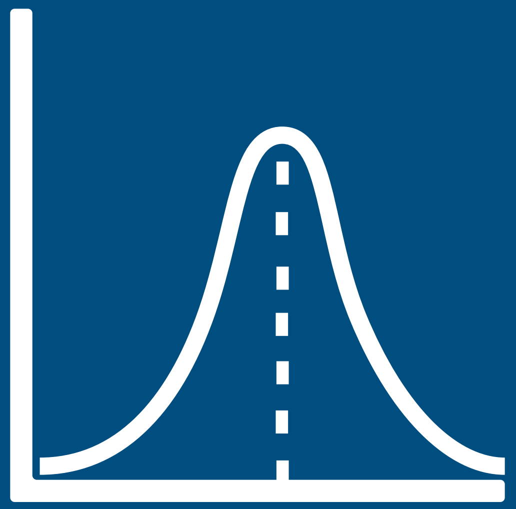
- Note that the lower limit a **cannot** be determined if $n_{obs} = 0$.
  - Inverse function does not exist!
  - By construction α is always equal to 1 for any a

$$\alpha = \sum_{n=n_{obs}}^{\infty} f(n;a) = 1 - \sum_{n=0}^{n_{obs}-1} f(n;a)$$

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

n_{obs}	lower limit a			upper limit b		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0	–	–	–	2.30	3.00	4.61

Confidence intervals using likelihood or LS



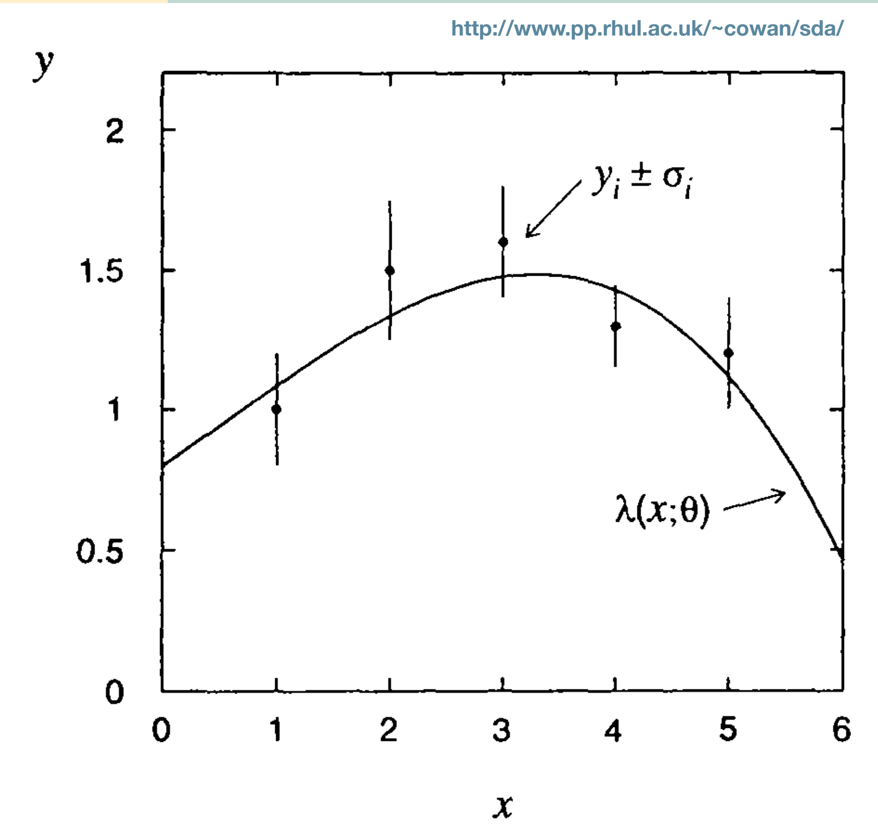
Approximations are sometimes best and surprisingly hard to break

Recall the Method of LS (L05,S34-39)

Review

Joint PDF is the product of N Gaussians

$$g(y_1, \ldots, y_N; \lambda_1, \ldots, \lambda_N, \sigma_1^2, \ldots, \sigma_N^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(y_i - \lambda(x_i; \boldsymbol{\theta}))^2}{2\sigma_i^2} \right)$$



Take the log (and drop additive terms that do not depend on the parameters)

$$\log \mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^{N} \frac{\left(y_i - \lambda(x_i; \boldsymbol{\theta})\right)^2}{\sigma_i^2}$$

Maximize by finding the values of the parameters θ that minimize $\chi^2(\theta)$

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{\left(y_i - \lambda(x_i; \boldsymbol{\theta})\right)^2}{\sigma_i^2}$$

$$\log \mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2}\chi^2(\boldsymbol{\theta})$$

or

$$\mathscr{L}(\boldsymbol{\theta}) = \exp\left(-\frac{\chi^2(\boldsymbol{\theta})}{2} \right)$$

Confidence Intervals with likelihood or LS

- Even in the case of a non-Gaussian estimator, the confidence intervals can be determined with a simple approximate technique
  - This makes use of the likelihood function or equivalently with the χ^2 function where one has the relation (see last slide)

$$\mathscr{L}(\boldsymbol{\theta}) = \exp\left(-\frac{\chi^2(\boldsymbol{\theta})}{2}\right)$$

- Let's first consider a ML estimator $\hat{\theta}$ for a parameter θ in the large sample limit:
  - The PDF $g(\hat{\theta}; \theta)$ does become Gaussian centered around the true value θ with a standard deviation of $\sigma_{\hat{\theta}}$

$$g(\hat{\theta}; \theta) = \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp\left(\frac{-(\hat{\theta} - \theta)^2}{2\sigma_{\hat{\theta}}^2}\right),$$

Large sample limit

- One can **also** show that in the large sample limit, the likelihood function becomes Gaussian in form centered about the ML estimate

$$L(\theta) = L_{\max} \exp\left(\frac{-(\theta - \hat{\theta})^2}{2\sigma_{\hat{\theta}}^2} \right) .$$

  - As discussed in recap, RCF inequality becomes an equality in the large sample limit and one can obtain the standard deviation via

$$\log L(\hat{\theta} \pm N\sigma_{\hat{\theta}}) = \log L_{\max} - \frac{N^2}{2} .$$

← find decreases by $N^2/2$ from maximum value of ML to estimate $N\sigma_{\hat{\theta}}$

Recall: For a Gaussian distributed estimator $\hat{\theta}$, the 68.3% central CI can be constructed from the estimator and its estimated std. dev. $\hat{\sigma}_{\hat{\theta}}$ as $[a, b] = [\hat{\theta} - \hat{\sigma}_{\hat{\theta}}, \hat{\theta} + \hat{\sigma}_{\hat{\theta}}]$, i.e., for a CL of $1 - \gamma$

Thus, the 68.3% CI is given by the values of θ at which the $\log \mathscr{L}$ function decreases by $1/2$ from it's max. value.

Non-Gaussian limit

- In fact, it can be shown that even if the likelihood function is not a Gaussian function of the parameters, the central confidence interval $[a, b] = [\hat{\theta} - c, \hat{\theta} + d]$ can still be approximated by using

$$\log L(\hat{\theta}^{+d}_{-c}) = \log L_{\max} - \frac{N^2}{2},$$

  - Here $N = \Phi^{-1}(1 - \gamma/2)$ is the quantile of the standard Gaussian corresponding to the desired confidence level $1 - \gamma$

One of the most commonly used methods to determine the statistical uncertainty

- Now use $\mathscr{L}(\boldsymbol{\theta}) = \exp\left(-\frac{\chi^2(\boldsymbol{\theta})}{2}\right)$ (i.e., a χ^2 fit with Gaussian errors),

and the prescription becomes

$$\chi^2(\hat{\theta}^{+d}_{-c}) = \chi^2_{\min} + N^2.$$

The proof that these intervals approximate the classical CI discussed earlier is beyond the scope of this course

Warning:

The correspondence with the classical CI developed in slides 12-27 <u>is only exact in the large sample limit</u>

$$\log L(\hat{\theta}^{+d}_{-c}) = \log L_{\max} - \frac{N^2}{2},$$

Many statisticians recommend using the term 'Likelihood interval' for an interval obtained from the likelihood function

$$\chi^2(\hat{\theta}^{+d}_{-c}) = \chi^2_{\min} + N^2.$$

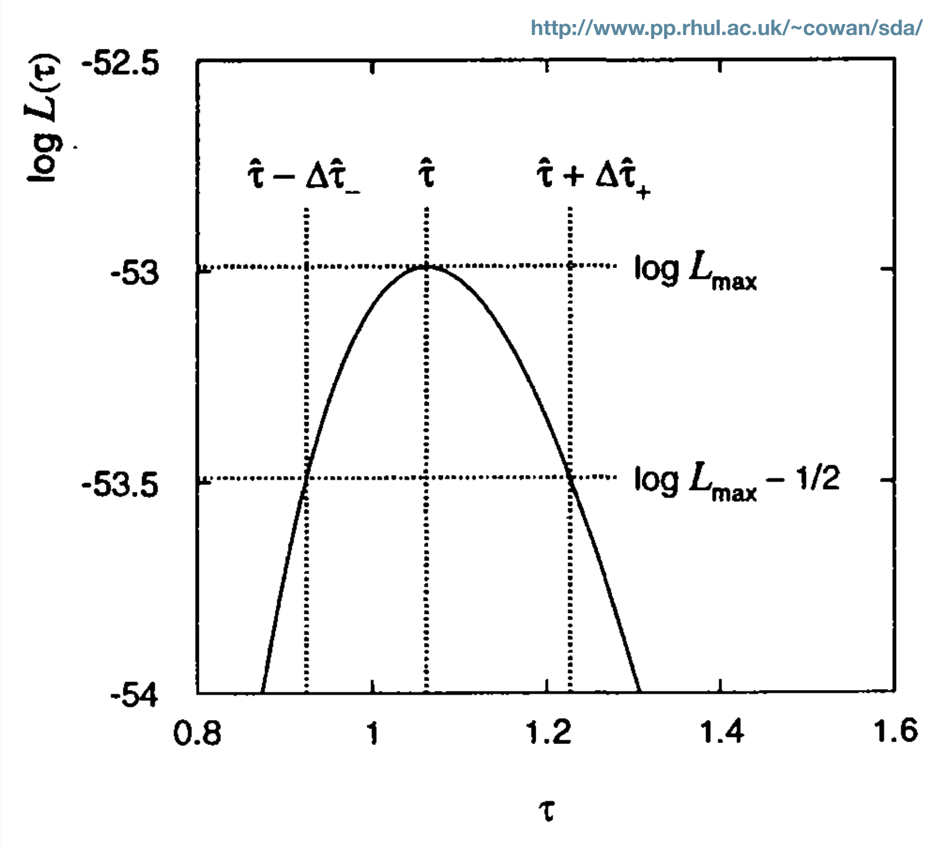
Interpret here as: an approximation to the classical CI, i.e., a random interval constructed so as to include the true parameter value with a given probability

A familiar example:

The estimator $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$ for the parameter τ of an exponential distribution $f(t;\tau) = \frac{1}{\tau}e^{-t/\tau}$

Recall Lecture 5 (sec. 6.2): The ML method was used to estimate τ given a sample of $n = 50$ measurements

- Reading off from the curve
  - $\Delta\hat{\tau}_- = 0.137$
  - $\Delta\hat{\tau}_+ = 0.165$
- Both reasonably close and we find
  - $\hat{\sigma}_{\hat{\tau}} \approx \Delta\hat{\tau}_- \approx \Delta\hat{\tau}_+ \approx 0.15$
- **In Lecture 5, we said:** we'll make a reinterpretation of the interval $[\hat{\tau} - \hat{\sigma}_{\hat{\tau}}, \hat{\tau} + \hat{\sigma}_{\hat{\tau}}]$ as an approximation of the **68.3% central confidence interval**

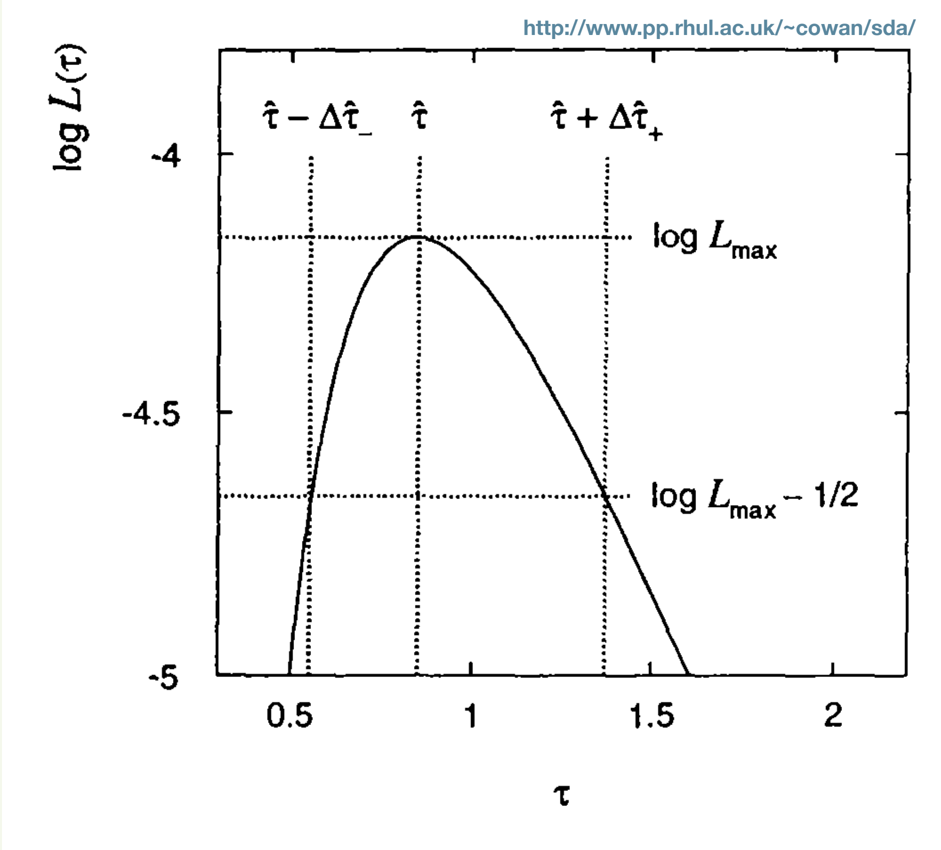


Now consider $n = 5$ measurements

- $\log \mathscr{L}(\boldsymbol{\theta})$ is less parabolic
  - so the half-width of the interval determined by $\log \mathscr{L}_{\max} - 1/2$ is not what we want to use to estimate the std. dev.
  - Better to use the CI to communicate the statistical uncertainty, since one then knows the probability that the interval covers the true parameter value.

$$\log \mathscr{L}(\tau) = \log \mathscr{L}_{\max} - 1/2$$

$$\hat{\tau} = 0.85^{+0.52}_{-0.30}$$

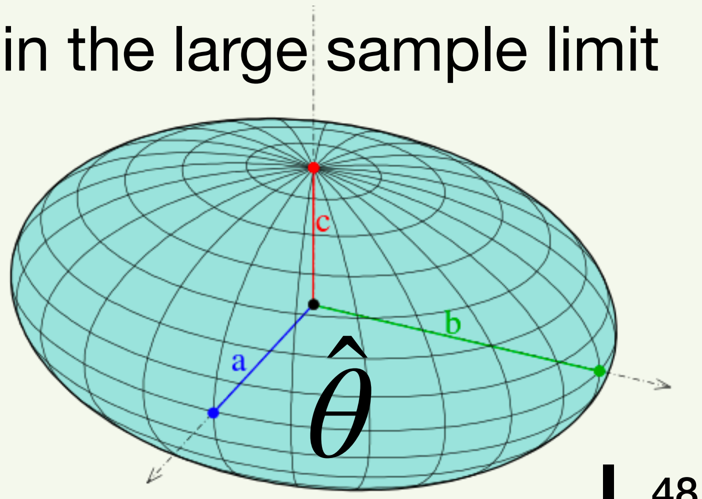


Multi-dimensional CL

Multi-dimensional confidence regions

- In 1D, we constructed a confidence interval $[a, b]$, to have a certain probability $1 - \gamma$ of containing a (true) parameter θ.
  - In order to generalize this to the case of n parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$, one might attempt to find an n-dimensional confidence interval $[\boldsymbol{a}, \boldsymbol{b}]$ constructed so as to have a given probability that $a_i < \theta_i < b_i$, simultaneously for all i.

    This turns out to be computationally difficult, not uniquely defined, and is thus rarely done.
  - It is nevertheless quite simple to construct a **confidence region** in parameter space such that the true parameter $\boldsymbol{\theta}$ **is contained** within the region with a given probability.
    - This region will not have the form $a_i < \theta_i < b_i$ with $i = 1,2, \ldots, n$, but will be more complicated, approaching an n-dimensional hyper-ellipsoid in the large sample limit



Multi-dimensional confidence regions

- As in the single-parameter case, one makes use of the fact that both the joint PDF for the estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$ as well as the likelihood function become **Gaussian** in **the large sample limit**. That is, the joint PDF of $\hat{\boldsymbol{\theta}}$ becomes:

$$g(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left[-\tfrac{1}{2} Q(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})\right],$$

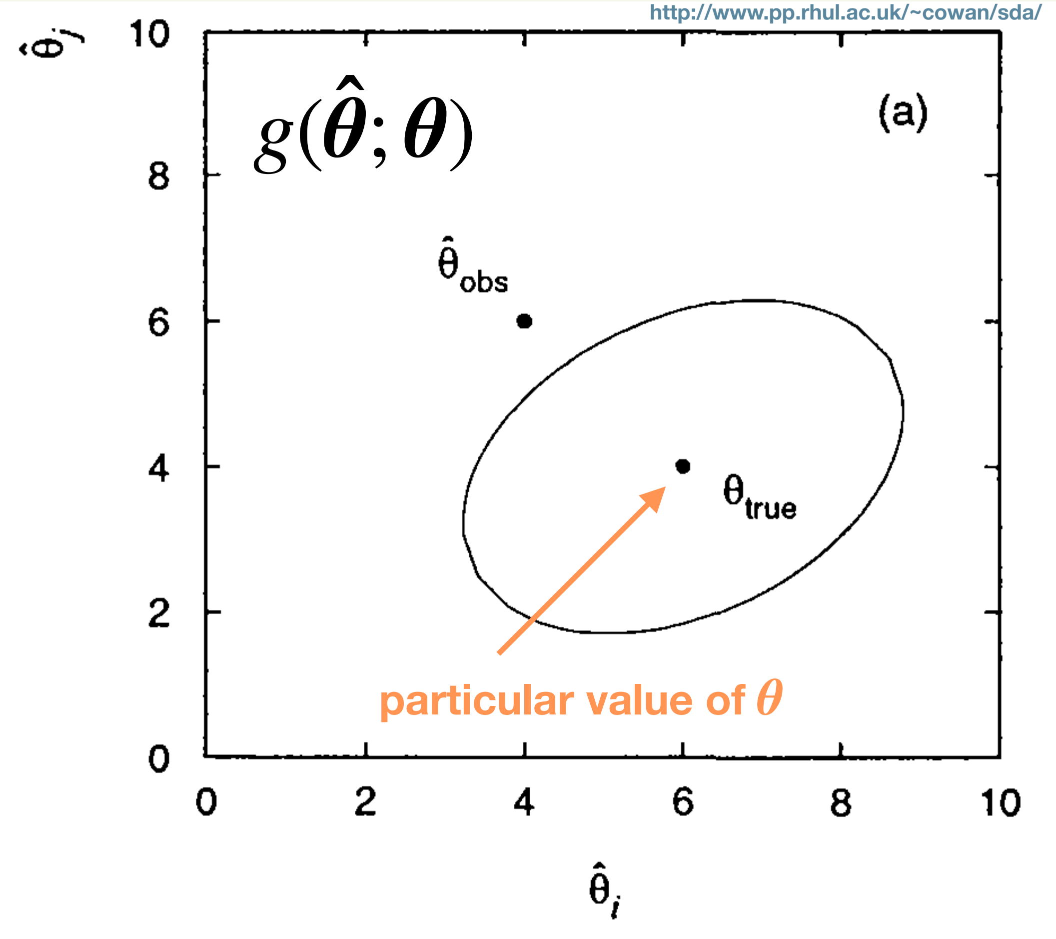
$$Q(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T V^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

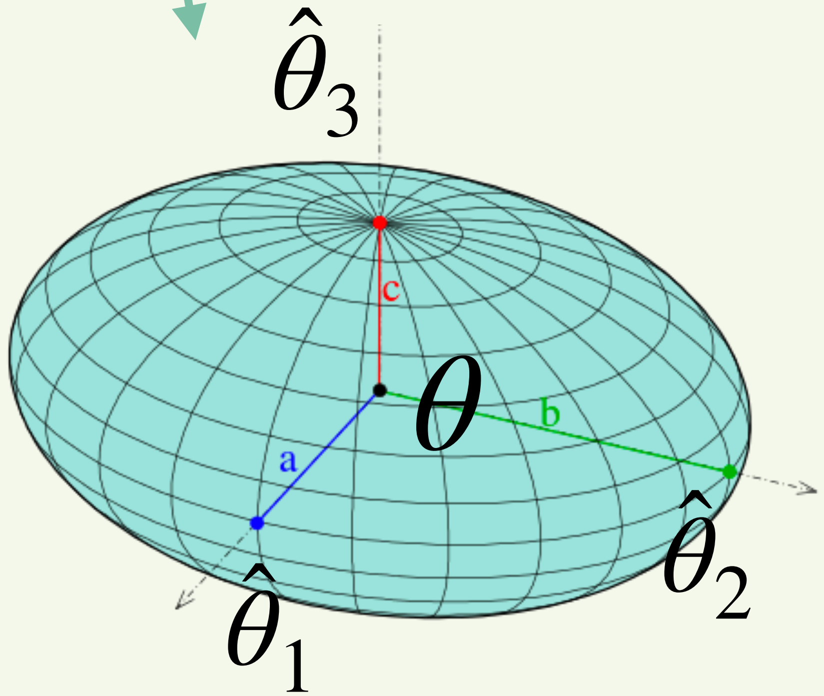
  - Here V^{-1} is the inverse covariance matrix and T refers to transposed.

Multi-dimensional confidence regions

$$Q(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T V^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

- Contours of constant $g(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta})$ correspond to constant $Q(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta})$
  - These are ellipses (or for more than 2D, hyper-ellipsoids) in $\hat{\theta}$ - space centered around the true parameters



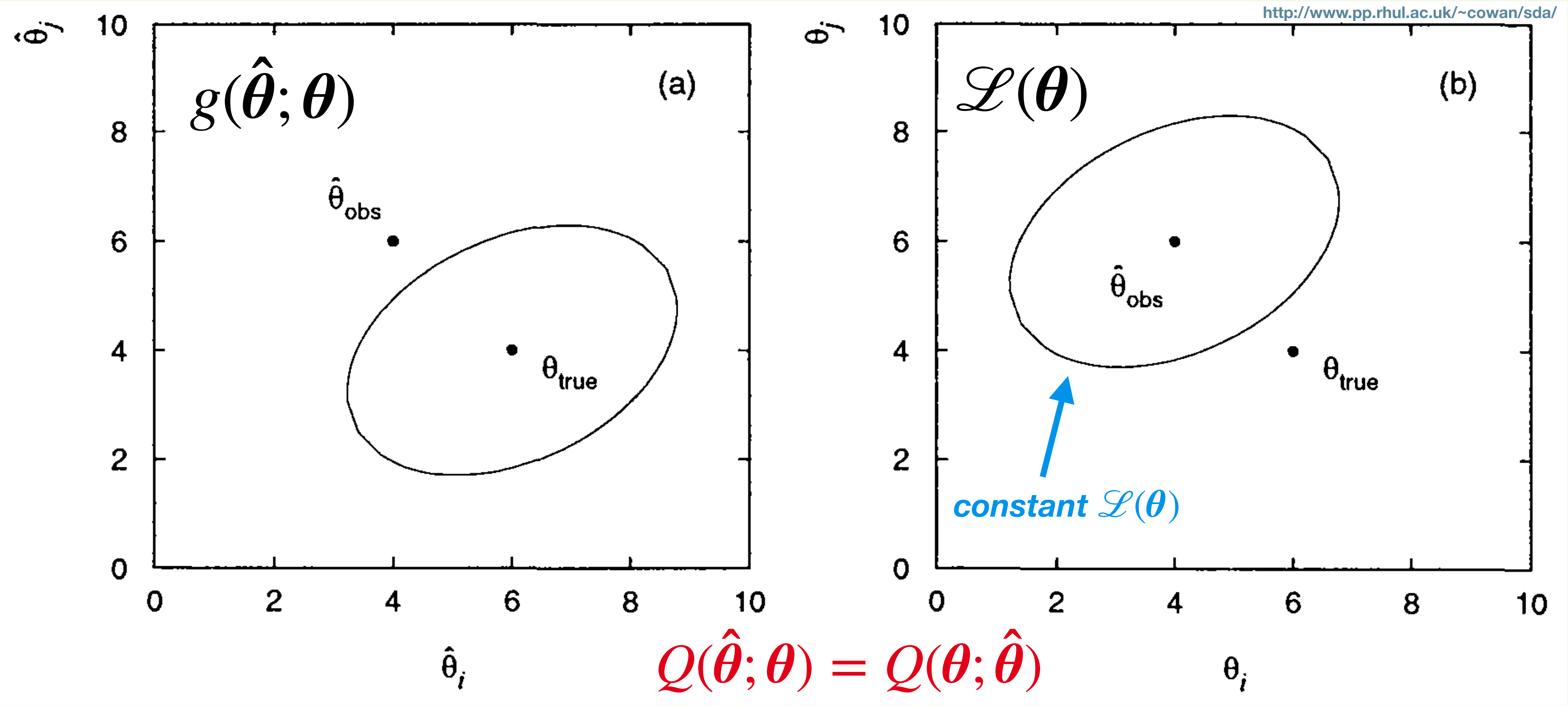


Multi-dimensional confidence regions

- Also, as in the one-dimensional case, one can show that the likelihood function $\mathscr{L}(\boldsymbol{\theta})$ takes on a Gaussian form centered about the ML estimate $\hat{\boldsymbol{\theta}}$

$$L(\boldsymbol{\theta}) = L_{\max} \exp\left[-\tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T V^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right] = L_{\max} \exp\left[-\tfrac{1}{2}\, Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})\right].$$

  - The function Q is here regarded as a function of the parameters $\boldsymbol{\theta}$ which has its maximum at the estimates $\hat{\boldsymbol{\theta}}$. Since Q is symmetric in $\hat{\boldsymbol{\theta}} \leftrightarrow \boldsymbol{\theta}$, Q is identical in $\mathscr{L}(\boldsymbol{\theta})$ & $g(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})$



Multi-dimensional confidence regions

- If $\hat{\boldsymbol{\theta}}$ is described by an n-dimensional Gaussian PDF $g(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})$, then the quantity $Q(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})$ is distributed according to a χ^2 with n *degrees of freedom*
  - The statement that $Q(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})$ is less than some value Q_γ

    (i.e., that the estimate $\hat{\boldsymbol{\theta}}$ is within a certain distance of the true value $\boldsymbol{\theta}$),
  - implies $Q(\hat{\boldsymbol{\theta}};\boldsymbol{\theta}) < Q_\gamma$

    (i.e., that the true value $\boldsymbol{\theta}$ is within the same certain distance of the estimate $\hat{\boldsymbol{\theta}}$).
  - **The two events therefore have the same probability:**

$$P(Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}) \le Q_\gamma) = \int_0^{Q_\gamma} f(z;n)\,dz, \;= 1-\gamma.$$

Chosen to correspond to a given probability content

This region of $\boldsymbol{\theta}$-space is called the confidence region with CL $1-\gamma$

χ^2 distribution for n degrees of freedom

Quantile of the order $1-\gamma$:

$$Q_\gamma = F^{-1}(1-\gamma;n)$$

Multi-dimensional confidence regions

- The region in $\boldsymbol{\theta}$-space defined by $Q(\hat{\boldsymbol{\theta}};\widehat{\boldsymbol{\theta}}) \leq Q_\gamma$ is called a **confidence region** with a **confidence level** of $\mathbf{1-\gamma}$
  - For a likelihood function of Gaussian form, it can be constructed by finding the values of $\boldsymbol{\theta}$ at which the log-likelihood function decreases by $Q_\gamma/2$ from its maximum value

$$\log L(\boldsymbol{\theta}) = \log L_{\max} - \frac{Q_\gamma}{2}.$$

  - Analogous to what was discussed before, one can still use the same prescription even if the likelihood function is not Gaussian
    - The **coverage** is then **only approximative** but in many use cases still adequate ↔ **need to check coverage using Monte Carlo**

Quantiles

- Quantiles for the χ^2 distribution $Q_\gamma = F^{-1}(1-\gamma; n)$ for several confidence levels $1-\gamma$ and $n = 1,2,3,4$ parameters are given below, as well as confidence levels for various values of the quantile Q_γ.

Q_γ	$1-\gamma$				
	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
1.0	0.683	0.393	0.199	0.090	0.037
2.0	0.843	0.632	0.428	0.264	0.151
4.0	0.954	0.865	0.739	0.594	0.451
9.0	0.997	0.989	0.971	0.939	0.891

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

$1-\gamma$	Q_γ				
	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
0.683	1.00	2.30	3.53	4.72	5.89
0.90	2.71	4.61	6.25	7.78	9.24
0.95	3.84	5.99	7.82	9.49	11.1
0.99	6.63	9.21	11.3	13.3	15.1

- **Note** that for **increasing** n the **confidence level** for a **given** Q_γ **decreases.**

  E.g. for $n=1$ $Q_\gamma = 1$ implies $1-\gamma = 0.683$. But $n=2$ $Q_\gamma = 1$ gives a confidence level of only 0.393

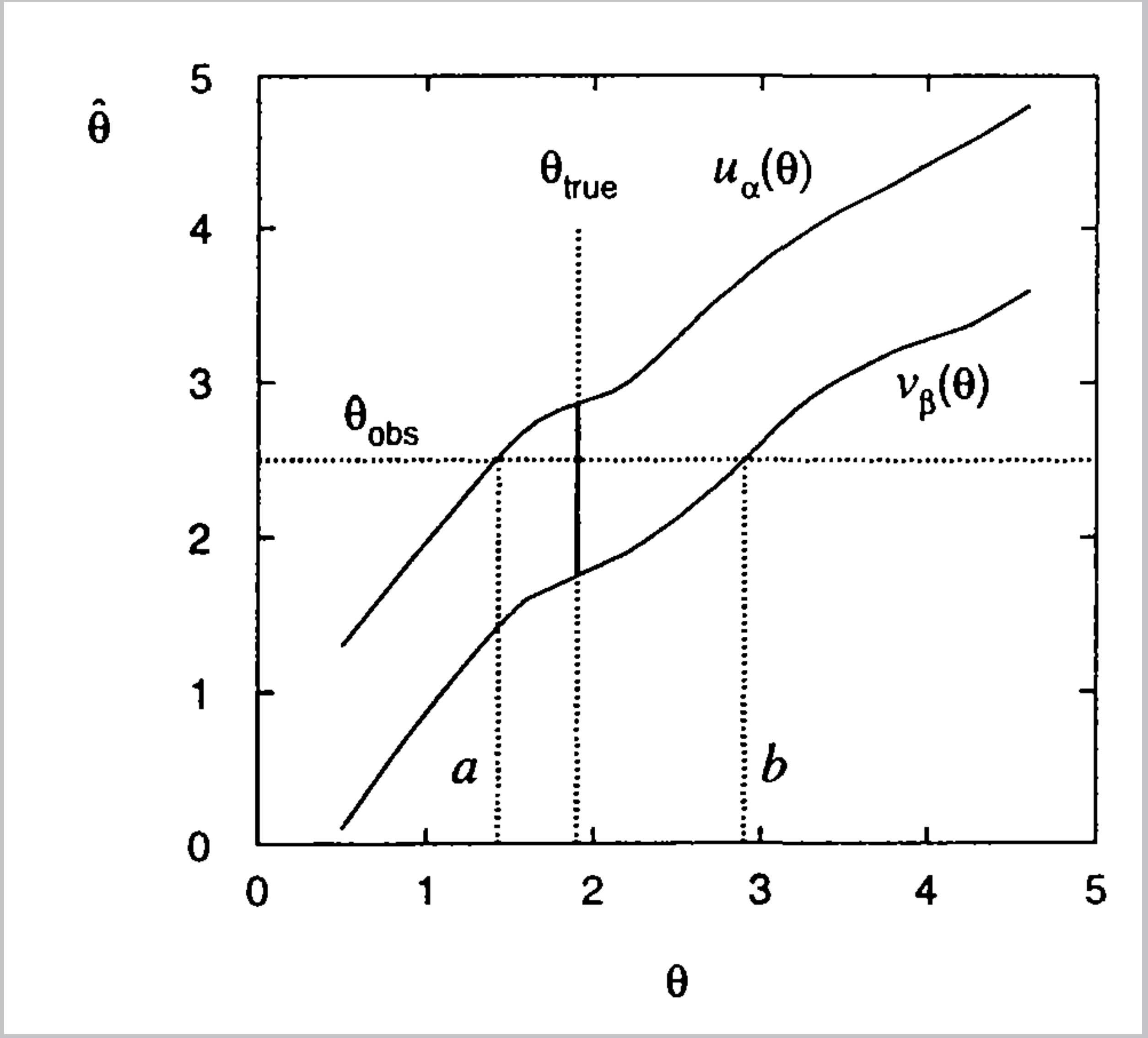
For next time

- Required reading
  - Cowan textbook: chapter 9 (through 9.7)

Quiz Time: 7th Round

Neyman Belt

1) Using the figure, explain how a confidence interval is constructed for an observation $\hat{\theta}_{\text{obs}}$



Poisson RV without background

Derive starting from

$$\alpha = P(\hat{\nu} \geq \hat{\nu}_{\text{obs}}) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{a^n}{n!} e^{-a}\,, \tag{1}$$

$$\beta = P(\hat{\nu} \leq \hat{\nu}_{\text{obs}}) = \sum_{n=0}^{n_{\text{obs}}} \frac{b^n}{n!} e^{-b} \tag{2}$$

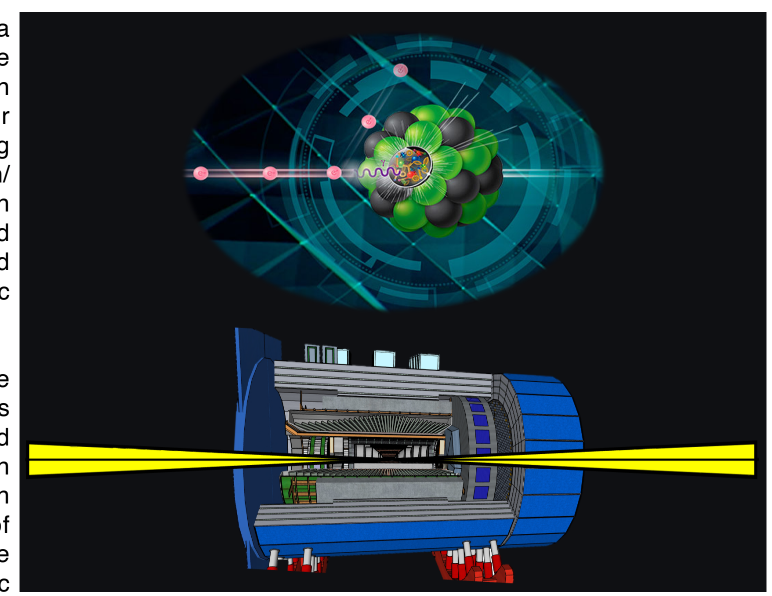
the Poisson upper-limit for $\hat{\nu} = n_{\text{obs}} = 0$ and $\alpha = \beta = 0.05$. Why is there no lower-limit?

KCETA Colloquium

From QCD to Visible Matter: An Insight into the U.S. Eletron-Ion Collider

Thursday, June 15, 2023
Kleiner Hörsaal A (CS) 15:45 - 17:00

Professor Or Hen
(Massachusetts Institute of Technology)



Recently the United States greenlit the construction of a revolutionary Electron-Ion Collider (EIC) at the Brookhaven National Lab. This once-in-a-generation $2.4 billion investment is set to propel our understanding of subatomic matter by generating unmatched high-current polarized electron and proton/ion beams that will interact at two distinct collision points. These interactions will be meticulously analyzed by cutting-edge detectors to uncover unprecedented insights into the formation and properties of subatomic matter.

The EIC's research program is primed to address some of the most profound questions in quantum physics encompassing the emergence of nucleon spin and mass the role of Quantum Chromodynamics (QCD) in nuclear interactions and its influence on bound nucleon structure and the three-dimensional structure of nucleons and nuclei. It will further delve into the uncharted territories of low-temperature dense gluonic matter properties and the quest for physics beyond the confines of the standard model.

In this talk I will elaborate on how the EIC will serve as an invaluable tool in addressing these perplexing questions. Additionally I will present an overview of the ePIC detector currently under design and construction by an international collaboration of scientists from over 160 institutions promising to redefine our understanding of the subatomic world.

Please note:
The colloquium will also be live-streamed to B402 SR224 (CN).

KIT Center Elementary Particle and Astroparticle Physics (KCETA)
www.kceta.kit.edu

Karlsruher Institut für Technologie

Bibliography

- Part of the material presented in this lecture is taken from the following sources. See the active links (when available) for a complete reference
  - **Statistical Data Analysis** textbook by G. Cowan (U. London): all figures & equations with white background