Statistical Models and Inference - Part III

Alberto Garfagnini

Università di Padova

AA 2020/2021 - Stat Lect. 7

Bayesian inference for a Bernoulli process

- we used Bayes' theorem

$$P\left(p \,\middle|\, \{y_j\}\ M\right) \propto f\left(\{y_j\} \,\middle|\, p\ M\right) \times g\left(p \,\middle|\, M\right)$$

- the likelihood of our data follows a binomial distribution

$$f(y \,|\, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

- multiplying a binomial likelihood times a beta prior ➜ new beta posterior distribution
- the beta distribution is the conjugate family for the binomial probability inference
  - $\mathtt{Beta}(1, 1)$ gives a uniform distribution
  - $\mathtt{Beta}(\frac{1}{2}, \frac{1}{2})$ gives a Jeffrey's prior: a distribution that is invariant under any continous transformation of the parameter
- the posterior distribution summarizes our belief about the parameter after having seen the data
- it takes into account our prior belief (the prior distribution) and the data (likelihood).
- we may want to determine an interval that has a high probability of containing the parameter. These are known as Bayesian credible intervals and are somewhat analogous to confidence intervals. But, they have the direct probability interpretation that confidence intervals lack

Bayesian inference for a Poisson process

- useful for counting the occurrences of rare events that happen at a constant rate both in time or in space
- example: number of accidents at a street crossing over a month
- the form of our Bayes' theorem is

$$P\left(\mu \,\middle|\, \{x_j\}\, M\right) \propto f\left(\{x_j\} \,\middle|\, \mu\, M\right) \times g\left(\mu \,\middle|\, M\right)$$

- $\{x_j\}$ indicates our measurement data set
- the parameter μ can assume any positive value ➜ use a continuous prior definite on positive values
- the scale factor is given by the evidence

  The normalized posterior is

$$P\left(\mu \,\middle|\, \{x_j\}\, M\right) = \frac{f\left(\{x_j\} \,\middle|\, \mu\, M\right) \times g\left(\mu \,\middle|\, M\right)}{\int f\left(\{x_j\} \,\middle|\, \mu\, M\right) \times g\left(\mu \,\middle|\, M\right) d\mu}$$

Likelihood for a Poisson process

- the likelihood for a single measurement of a Poisson process is

$$f(x \,|\, \mu) = \frac{\mu^x e^{-\mu}}{x!}$$

- with $\mu > 0$ and $x = 0, 1, \ldots$
- in case of multiple independent measurements, the likelihood becomes

$$f(\{x_j\} \,|\, \mu) = \prod_{j=1}^{n} f(x_j \,|\, \mu) \quad \propto \quad \mu^{\sum x_j} \times e^{(-n\mu)}$$

- the function looks similar to the `Gamma` distribution function:

$$\texttt{Gamma}(x \,|\, \alpha, \lambda) = k\, x^{\alpha-1}\, e^{-\lambda x}$$

  with

$$k = \frac{\lambda^\alpha}{\Gamma(\alpha)}$$

- comparing with our case: $\alpha = \sum x_j + 1$ and $\lambda = n$

Posterior for a Poisson process (1)

- according to the background knowledge of the researcher, we can have different prior distribution functions:
  - a positive uniform prior
  - a Jeffrey's prior, which is invariant under any continuous transformation of the parameter
  - a Gamma prior, which is the conjugate family for the Poisson inference

Uniform prior

- it is used when there is no idea on what the μ parameter value could be

$$g(\mu) = 1 \quad \text{for} \quad \mu > 0$$

- it is also called an improper prior, since the integral on all the possible value of the parameter diverges
- the posterior becomes:

$$\begin{aligned} P(\mu \,|\, \{x_j\}) &\propto f(\{x_j\} \,|\, \mu) \times g(\mu) \\ &\propto \mu^{\sum x_j} \quad e^{-n\mu} \end{aligned}$$

- it's a $\mathtt{Gamma}(\alpha, \lambda)$ function with $\alpha = \sum x_j + 1$ and $\lambda = n$

Posterior for a Poisson process (2)

Jeffrey's prior

- it is a pior whichis invariant under any continuous transformation of the parameter

$$g(\mu) \propto \frac{1}{\sqrt{\mu}} \quad \text{for} \quad \mu > 0$$

- it is an improper prior ➜ its integral over the whole parameter range in infinite
- combining with the likelihood, we get

$$\begin{aligned} P(\mu \,|\, \{x_j\}) &\propto f(\{x_j\} \,|\, \mu) \times g(\mu) \\ &\propto \mu^{\sum x_j} \quad e^{-n\mu} \times \frac{1}{\sqrt{\mu}} \\ &\propto \mu^{\sum x_j - 1/2} \quad e^{-n\mu} \end{aligned}$$

- it's again a $\mathtt{Gamma}(\alpha, \lambda)$ function with $\alpha = \sum x_j + \frac{1}{2}$ and $\lambda = n$

Posterior for a Poisson process (3)

Conjugate family prior

- the conjugate family of functions for the Poisson process with parameter μ will have the same form of the likelihood

$$
\begin{aligned}
g(\mu) &\propto e^{-k\,\mu} \quad e^{\log \mu \times r} \\
&\propto e^{-k\,\mu} \quad \mu^{r}
\end{aligned}
$$

- a distribution having this shape is the $\mathtt{Gamma}(\alpha, \lambda)$ function with $\alpha - 1 = r$ and $\lambda = k$
- the normalization scale factor is $\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}$

Posterior for a single observation

- using a $\mathtt{Gamma}(\alpha, \lambda)$ prior

$$
\begin{aligned}
P(\mu \,|\, y) &\propto f(y \,|\, \mu) \times g(\mu) \\
&\propto \frac{\mu^{y} e^{-\mu}}{y!} \times \frac{\lambda^{\alpha} \quad \mu^{\alpha-1} \quad e^{-\lambda\mu}}{\Gamma(\alpha)} \\
&\propto \mu^{\alpha-1+y} \quad e^{-(\lambda+1)\mu}
\end{aligned}
$$

- it's a $\mathtt{Gamma}(\alpha', \lambda')$ with $\alpha' = \alpha + y$ and $\lambda' = \lambda + 1$

Posterior for a Poisson process (4)

Posterior for multiple observations

- we have a set of n observations $\{y_j\}$

- assume a $\mathtt{Gamma}(\alpha, \lambda)$ prior
  - ➜ a uniform prior, $g(\mu) = 1$, has the form of $\mathtt{Gamma}(1, 0)$
  - ➜ the Jeffrey's prior for Poisson, $g(\mu^{-1/2})$ is equivalent to $\mathtt{Gamma}(\frac{1}{2}, 0)$
    they are both considered as a limiting case of $\mathtt{Gamma}(\alpha, \lambda)$, with $\lambda \to 0$

- start with the first observation and evaluate the posterior distribution
- repeat the updating after each observation, using the posterior from the j-th observation as the prior for the observation j+1

- we end up with a $\mathtt{Gamma}(\alpha', \lambda')$ posterior where

$$
\alpha' = \alpha + \sum y \quad \text{and} \quad \lambda' = \lambda + n
$$

- the expected value and variance of the posterior are

$$
E[\mu \,|\, y] = \frac{\alpha'}{\lambda'} \quad \text{and} \quad \mathrm{Var}[\mu \,|\, y] = \frac{\alpha'}{\lambda'^{2}}
$$

How to choose the conjugate prior

- the $\texttt{Gamma}(\alpha, \lambda)$ family of distributions is the conjugate family for the inference of μ parameter from a Poisson distribution
- the advantage of using a conjugate prior is that the posterior will be from the same family and can be found with simple updating rules
- to determine the parameters of the $\texttt{Gamma}(\alpha, \lambda)$ prior that matches our belief, try to summarize your belief into a prior mean m and a prior standard deviation s
- since

$$m = \frac{\alpha}{\lambda} \quad \text{and} \quad s^2 = \frac{\alpha}{\lambda^2}$$

- we get, by inversion:

$$\lambda = \frac{m}{s^2} \quad \text{and} \quad \alpha = \left(\frac{m}{s}\right)^2$$

Exercise

- the weekly number of traffic accidents on a highway between two towns follows a Poisson distribution
- four students are going to count the number of traffic accidents for the next eight weeks

1 Student 1 has no prior information. Therefore she will assume all possible values for μ are equally likely

➜ positive uniform prior, $g(\mu) = 1$

2 Student 2 has no prior information, either. But she wants her prior to be invariant if the parameter is multiplied by a constant

➜ Jeffrey's prior, $g(\mu) = \mu^{-1/2}$

3 Student 3 believes the prior mean should be 2.5 with a standard deviation of 1

➜ Gamma prior, $\texttt{Gamma}(\alpha = 2.5, \lambda = 6.25)$

4 Student 4 thinks his prior has a trapezoidal shape. He draws the prior function by interpolating the values with the following weights

Accidents	0	2	4	8	10
Weight	0	2	2	0	0

Exercise : solution

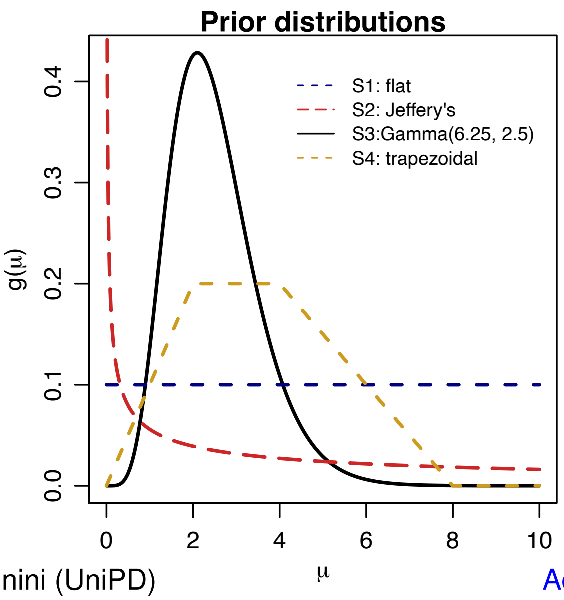
- the 8 weeks measurements bring the following number of accidents per week:
  `3, 2, 0, 8, 2, 4, 6, 1`
- let's evaluate the posteriors:

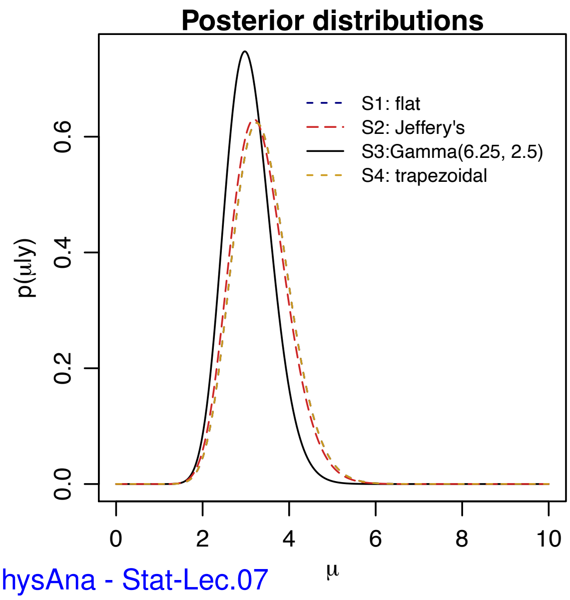
□→ Student 1 : $\texttt{Gamma}(\alpha', \lambda')$ where $\alpha' = \sum y_j + 1 = 27$ and $\lambda' = n = 8$

◇→ Student 2 : $\texttt{Gamma}(\alpha, \lambda')$ where $\alpha' = \sum y_j + \frac{1}{2} = 26.5$ and $\lambda' = n = 8$

○→ Student 3 : $\texttt{Gamma}(\alpha', \lambda')$ where $\alpha' = \sum y_j + \alpha = 32.25$ and $\lambda' = \lambda + n = 10.5$

⊡⇒ Student 4 : numerical integration





Example : quantitative results

- the posterior distribution is the complete inference in Bayesian modeling and it allows to extract all the possible parameter values
- three possible measure of the location of a distribution are:
  - the mean
  - the mode (i.e. the most probable value)
  - the median (i.e. the value for which $P(x \leq \text{med}) = P(x > \text{med}) = 0.5$)
- we summarize by extracting the numerical estimates for the four students

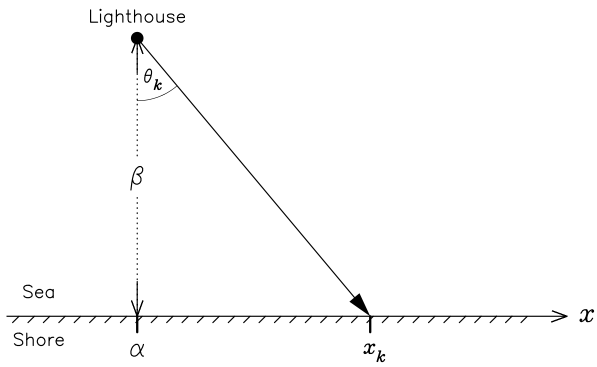
Student	S1	S2	S3	S4
Post	$\texttt{Gamma}(27, 8)$	$\texttt{Gamma}(26.5, 8)$	$\texttt{Gamma}(32.25, 10.5)$	numerical
Mean	3.37	3.31	3.07	3.35
Median	3.33	3.27	3.04	3.32
Mode	3.25	3.19	2.98	-
Std. Dev.	0.65	0.64	0.54	0.63

Credibility Interval 95%

low	2.22	2.17	2.10	2.22
high	4.76	4.69	4.22	4.67

The Lighthouse problem

- A lighthouse is located at a position α along the shore and at a distance β out at sea
- It emits a series of short highly collimated flashes at random intervals and at random angles
- we detect the pulses on the coast using photo-detectors; they record only the position x_k of the flash arrival on the coast, but not the angle of emission



- N flashes have been recorded at positions $\{x_k\}$ ➜ We want to estimate the position of the lighthouse
- it looks reasonable to assign a uniform Likelihood pdf on the azimuth angle θ_k

$$P\left(\theta_k \,\middle|\, \alpha, \beta\right) = \frac{1}{\pi}$$

- where θ_k is connected to α and β by the relation

$$x_k - \alpha = \beta \tan \theta_k$$

- we operate a change of variable on the pdf

$$P(X|M) = P(Y|M)\left|\frac{dY}{dX}\right|$$

The Lighthouse problem

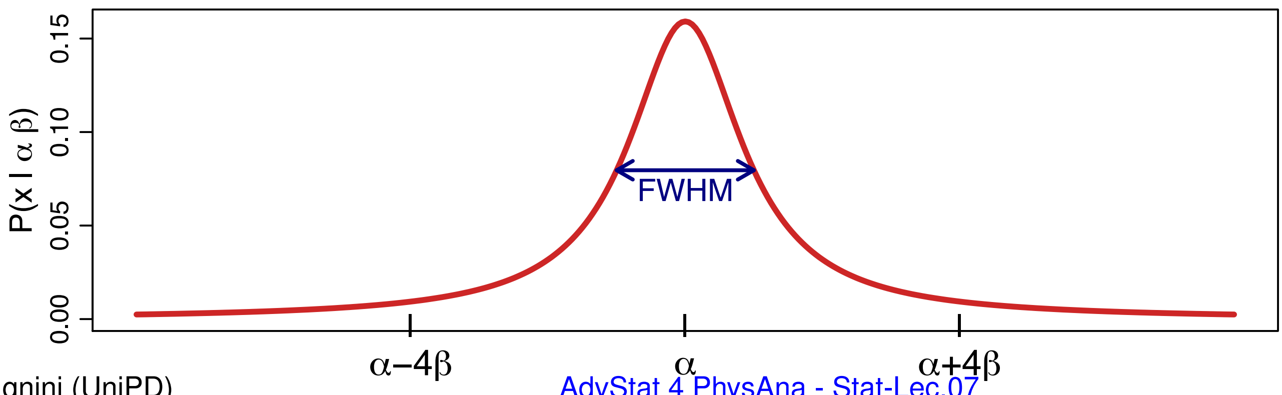
- applying the transformation

$$\begin{aligned} x &= \beta \tan\theta + \alpha \\ dx &= \beta \frac{1}{\cos^2\theta} d\theta = \beta(1+\tan^2\theta) d\theta \\ \left|\frac{dx}{d\theta}\right| &= \beta(1+\tan^2\theta) = \beta\left[1 + \frac{(x-\alpha)^2}{\beta^2}\right] = \frac{\beta^2 + (x-\alpha)^2}{\beta} \end{aligned}$$

- we get

$$P(x \,|\, \alpha, \beta) = P(\theta|\, \alpha, \beta)\left|\frac{d\theta}{dx}\right| = \frac{1}{\pi}\frac{\beta}{\beta^2 + (x-\alpha)^2}$$

- we have obtained a Cauchy distribution, which is symmetric about the maximum α and with a FWHM of 2β



The Lighthouse problem

- we assume the distance (i.e. β) is known and the only missing parameter is α
- from Bayes' theorem

$$P\left(\alpha \mid D, \beta\right) \propto P\left(D \mid \alpha, \beta\right) \times P\left(\alpha \mid \beta\right)$$

- since β tells us nothing about α, we assume a uniform prior

$$P(\alpha \mid \beta) = P(\alpha) = \begin{cases} \dfrac{1}{\alpha_{max} - \alpha_{min}} & \text{for } x \in [\alpha_{min}, \alpha_{max}] \\ 0 & \text{otherwise} \end{cases}$$

- the recording of a signal at one photo-detector does not influence what we can infer about the position of another measurement (given the same location of the lighthouse)

➜ the Likelihood function is just the product of the probabilities for N individual detections

$$P\left(D \mid \alpha, \beta\right) = \prod P\left(x_j \mid \alpha, \beta\right)$$

- taking the natural logarithm of the Posterior probability function

$$L = \ln P\left(\alpha \mid D, \beta\right) = \text{const} - \sum \ln\left[\beta^2 + (x_j - \alpha)^2\right]$$

- where `const` includes all the terms not depending on α

The Lighthouse problem

- the best estimate $\alpha_\circ$ is given by the maximum of the posterior pdf

$$\left.\frac{dL}{d\alpha}\right|_{\alpha_\circ} = 2 \sum_j \frac{x_j - \alpha_\circ}{\beta^2 + (x_j - \alpha_\circ)^2} = 0$$

- an analytical solution is difficult to be done, but nothing stops us from evaluating it numerically

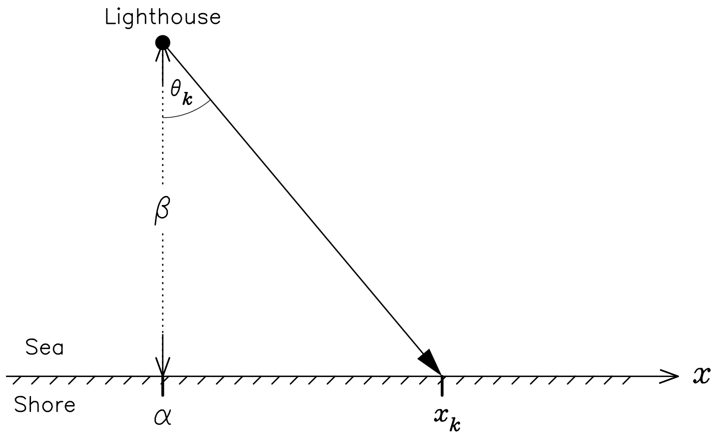
Homework

- write a small R program to evaluate the posterior distribution as a function of the collected data
- assume $\beta = 1$ km, $\alpha_{TRUE} = 1$ km and sample data in the range $x \in [-2 \text{ km}, +2 \text{ km}]$
- plot the posterior as a function of the number of collected data (assume $n = \{1, 2, 5, 10, 20, 50, 100\}$)

The Lighthouse problem

Homework

- write a small R program to evaluate the posterior distribution as a function of the collected data
- assume $\beta = 1$ km, $\alpha_{TRUE} = 1$ km and sample data in the range $x \in [-2 \text{ km}, +2 \text{ km}]$
- plot the posterior as a function of the number of collected data (assume $n = \{1, 2, 5, 10, 20, 50, 100\}$)



- our Swiss Army knife is always Bayes' theorem

$$P\left(\alpha \mid D, \beta\right) \propto P\left(D \mid \alpha, \beta\right) \times P\left(\alpha \mid \beta\right)$$

A solution to the Lighthouse problem

- instead of calculating the Posterior, it is better to evaluate the logarithm of the posterior

$$L = \log P\left(\alpha \mid \{x_k\}, \beta\right) = \text{const} - \log\left(1 + \left(\frac{x - \alpha}{\beta}\right)^2\right)$$

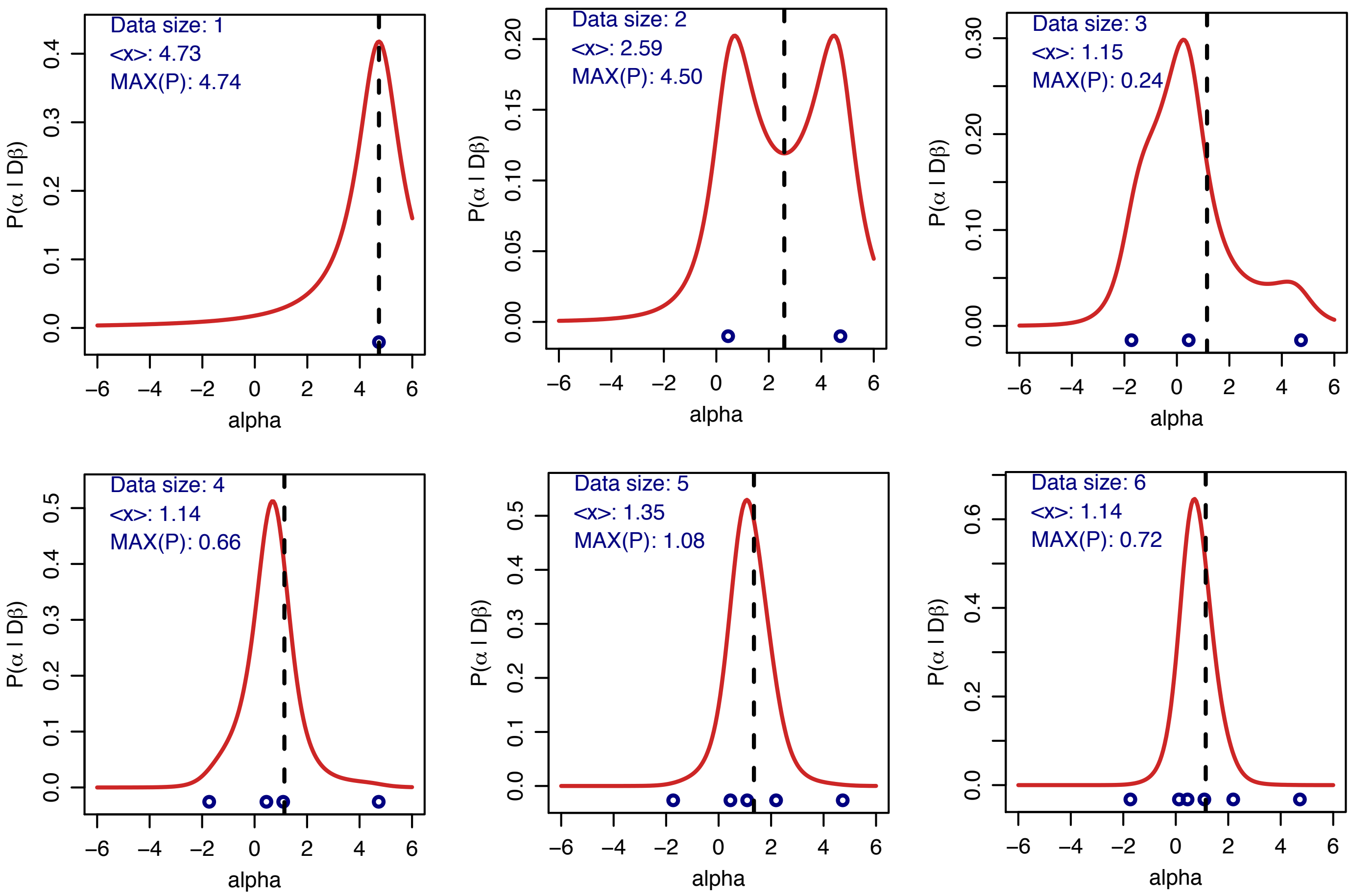
- and afterwards take the exponential

```
p.log.like <- function(a, data) {
  b <- 1
  logL <- 0.0
  for (x in data) {
    logL <- logL - log(1 + ((x-a)/b)^2)
  }
  return(logL)
}

n.sample <- 200
x.min <- -6; x.max <- +6
h <- (x.max - x.min)/n.sample
alpha <- seq(from=x.min, by=h, length.out=n.sample+1)
```

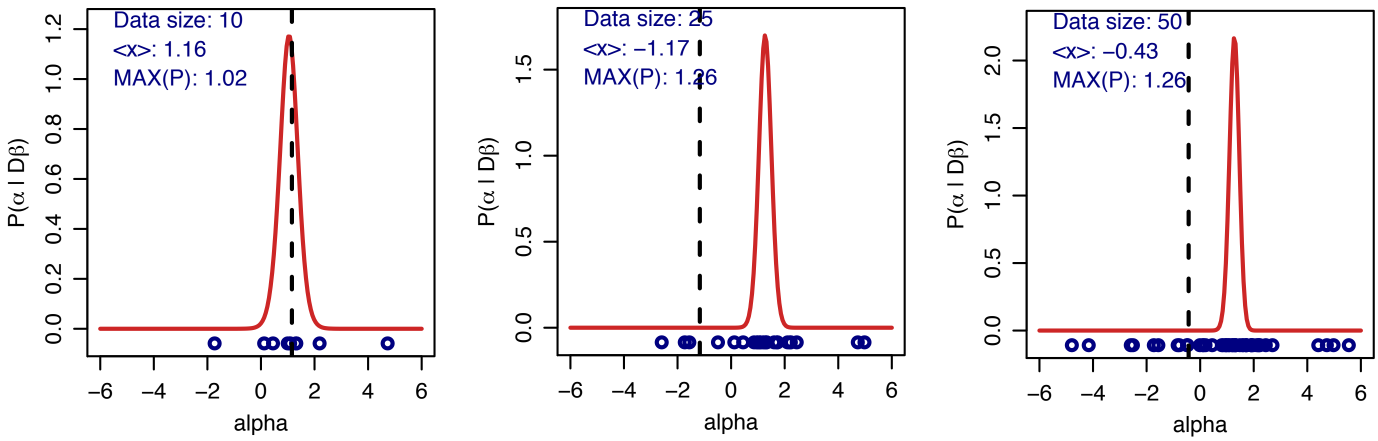
A solution to the Lighthouse problem

```
n.str <- readline("Enter␣data␣set␣dimension:␣")
n.plot <- as.numeric(unlist(strsplit(n.str, ",")))
dt <- data[1:n.plot]

# Get the LogLikelihood
y.log.star <- p.log.like(alpha, dt)
# - Find the maximum
index.max <- which.max(y.log.star)
alpha.max <- alpha[index.max]
cat(paste("Alpha␣max:␣", alpha.max, '\n'))

# - get the exponential and normalize the posterior
y.post.star <- exp(y.log.star)
y.post <- y.post.star/(h*sum(y.post.star))

plot(alpha, y.post, type='l', lwd=2, col='firebrick3')

dt.mean <- mean(dt)
abline(v=dt.mean, lty=2, lwd=2)

y.band <- (max(y.post) - min(y.post))*0.05
text(-6, max(y.post)+y.band, col='navy', lwd = 2, pos=4,
     paste("Data␣size:␣", n.plot, sep=''))
text(-6, max(y.post)-y.band, col='navy', lwd = 2, pos=4,
     sprintf("<x>:␣%.2f", dt.mean))
text(-6, max(y.post)-3*y.band, col='navy', lwd = 2, pos=4,
     sprintf("MAX(P):␣%.2f", alpha.max))
```


The Posterior of the Lighthouse problem



The Posterior of the Lighthouse problem



- the positions of the flashes are marked by the open circles
- the posterior is very broad for small data sets and can also be multimodal if the flashes locations are well separated
- already with $\sim$ 10 measurements the posterior becomes a well shaped-like Gaussian
- it becomes narrower as the data size increases (FWHM $\propto 1/\sqrt{N}$)

➜ a simple average of the measurement data gives us a wrong result

Signal Amplitude in presence of Background

- given a set of counts $\{N_k\}$, masured at values $\{x_k\}$ we want the best estimate of the amplitude of the signal peak and of the background staying below
- for instance, in case of a photon spectrum, we measure the number of photons in bins of wavelength or energy
- this number is proportional to the exposure (time of measurement) and to both signal and background amplitudes through the expression

$$S_k = \Delta t \left[A \exp\left(-\frac{(x_k - x_\circ)^2}{2w^2} \right) + B \right]$$

where Δt is the exposure time and $x_\circ$ and w are the centre and widh of the signal peak

- the number of expected photons is S_k, not generally an integer
- the number of observed photons, N, is an integer number and follows the Poisson distribution

$$P(N|S) = \frac{S^N e^{-S}}{N!}$$

- and this gives us the Likelihood of the data ($D = \{N_j\}$)

$$P(D \,|\, A, B, M) = \prod_j \frac{S_k^{N_k} e^{-S_k}}{N_k!}$$

Signal Amplitude in presence of Background

- the model has 5 parameters, but we assume that $x_\circ$, w and Δt are known
- we want to infer $P(A, B \mid D, M)$ from the data, where M identifies the model (the shape of the line and the values of the fixed parameters)
- we adopt a minimalistic prior: we only assume that A and B cannot be negative. Therefore the Prior $P(A, B \mid M)$ is constant when both A and B are positive, and zero otherwise
- the Posterior is

$$P\,(A, B \mid D, M) = \frac{1}{Z} \prod_j \frac{S_k^{N_k}\, e^{-S_k}}{N_k!}$$

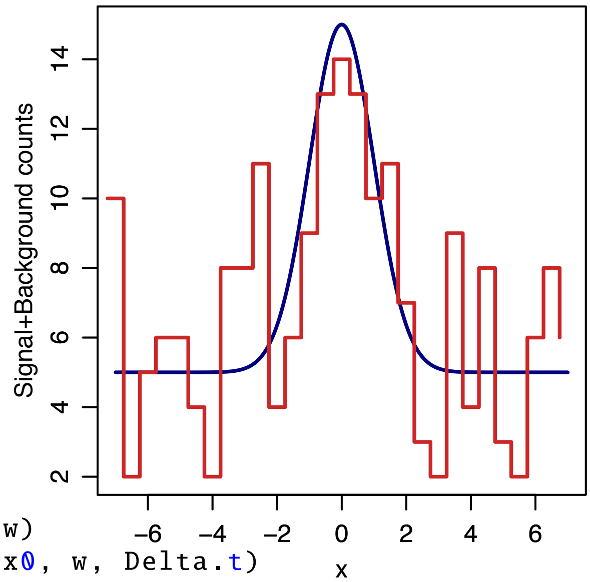
- and the log Posterior

$$L = \log P(A, B \mid D, M) = \text{const} + \sum [N_k \log S_k - S_k]$$

- where the constant term absorbs terms that do not depend on A or B
- the best estimates are given by values of A and B that maximize L

Signal+Background - data generation

- given the positions $\{x_k\}$, the number of expected photons is evaluated using the generative model

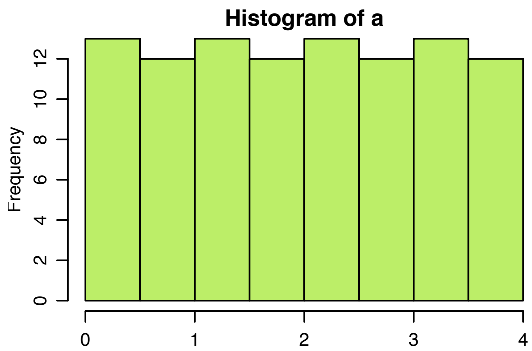


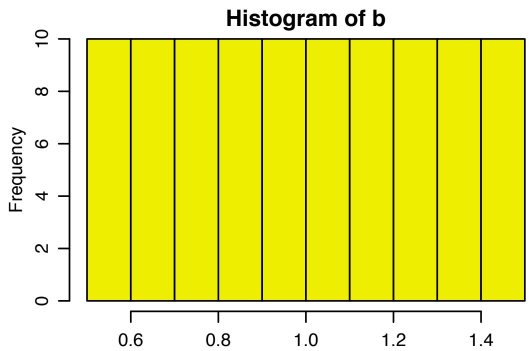

```r
# - Generative model
signal <- function(x, a, b, x0, w, t) {
    t * (a*exp(-(x-x0)^2/(2*w^2)) + b)
}

# Define model parameters
x0   <- 0      # Signal peak
w    <-1       # Signal width
A.true   <- 2 # Signal amplitude
B.true   <- 1 # Background amplitude
Delta.t  <- 5 # Exposure time

# - Generate the observed data
set.seed(205)
xdat   <- seq(from=-7*w, to=7*w, by=0.5*w)
s.true <- signal(xdat, A.true, B.true, x0, w, Delta.t)
ddat   <- rpois(length(s.true), s.true)

xplot <- seq(from=min(xdat), to=max(xdat), by=0.05*w)
splot <- signal(xplot, A.true, B.true, x0, w, Delta.t)
plot(xplot, splot,
     xlab="x", ylab="Signal+Background␣counts")
par(new=TRUE)
xdat.off <- xdat-0.25
plot(xdat.off, ddat, type='s',col='firebrick3',
     lwd=2,xlim=range(xplot), ylim=range(c(splot, ddat)))
```

Signal+Background - posterior calculation 1

- the posterior has a nonlinear dependence on the parameters
- we calculate it on a grid of values of $\{a_k, b_k\}$
- a regular grid of size $K \times K$, with $K = 100$ is used and the `contour()` function is used to plot lines of constant probability density

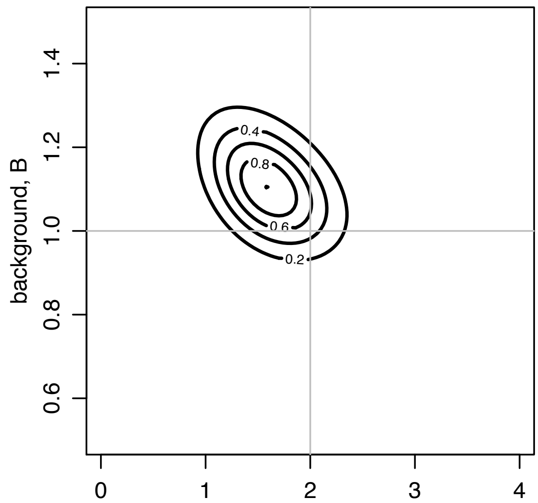
```
# - Sampling grid for computing posterior
alim   <- c(0.0, 4.0)
blim   <- c(0.5, 1.5)
Nsamp <- 100
uniGrid <- seq(from=1/(2*Nsamp),
to=1-1/(2*Nsamp), by=1/Nsamp)
delta_a <- diff(alim)/Nsamp
delta_b <- diff(blim)/Nsamp
a <- alim[1] + diff(alim)*uniGrid
b <- blim[1] + diff(blim)*uniGrid
```





Signal+Background - posterior calculation 2

```
# Log posterior
log.post <- function(d, x, a, b, x0, w, t) {
    if(a<0 || b <0) {return(-Inf)} # the effect of the prior
    sum(dpois(d, lambda=signal(x, a, b, x0, w, t), log=TRUE))
}

# Compute log unnormalized posterior, z = ln P^*(a,b|D), on a regular grid
z <- matrix(data=NA, nrow=length(a), ncol=length(b))
for(j in 1:length(a)) {
    for(k in 1:length(b)) {
        z[j,k] <- log.post(ddat, xdat, a[j], b[k], x0, w, Delta.t)
    }
}
z <- z - max(z) # set maximum to zero

# Plot unnormalized 2D posterior as contours.
contour(a, b, exp(z),
        nlevels = 5,
        labcex = 0.5,
        lwd = 2,
        xlab="amplitude,␣A",
        ylab="background,␣B")
abline(v=2,h=1,col="grey")
```



Marginal posterior pdf

- given our two parameters posterior pdf

$$P\left(A,B \,\middle|\, D\right) = \frac{1}{Z_{ab}} P\left(D \,\middle|\, A,B\right) P\left(A,B\right)$$

- we are interested in the posterior only for one parameter, for instance A ➜ we must *marginalize* (integrate) over B

$$P\left(A \,\middle|\, D\right) = \int P\left(A,B \,\middle|\, D\right) dB = \frac{1}{Z_{ab}} \int P\left(D \,\middle|\, A,B\right) P\left(A,B\right) dB$$

- if the Priors are independent, $P(AB) = P(A)P(B)$, marginalizing is like projecting the distribution along an axis

➜ marginalization is a powerful feature of probability analysis because it allows us to include parameters which are an essential part of the model, but which we may not actually be interested in

- We marginalize over them to get the posterior pdf for the parameters of interest

Signal+Background - marginalization

- the marginalization is performed on the grid, simply by summing over one of the two parameters

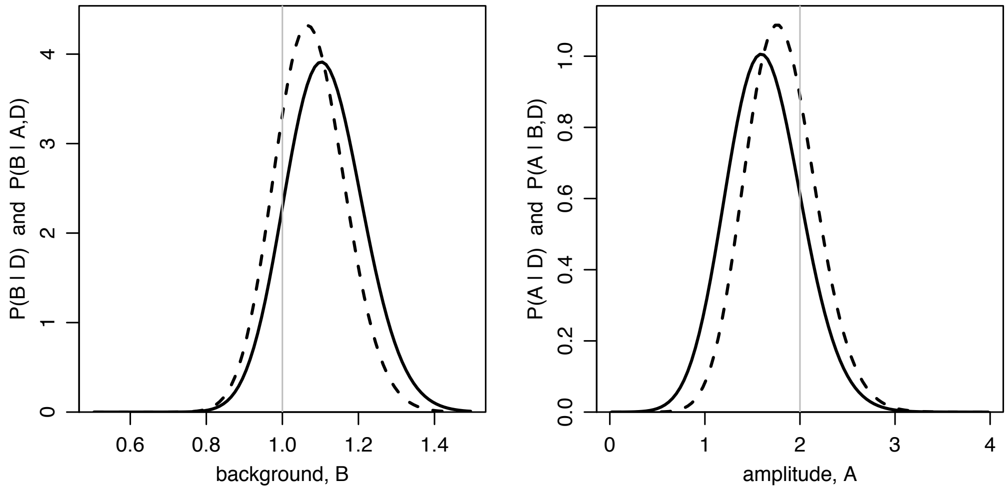
$$P\left(A_J \,\middle|\, D\right) \sim \Delta B \sum_{k=1}^{K} P\left(A_j, B_k \,\middle|\, D\right)$$

- where K is the grid size
- finally, the posterior is normalized in via the rectangle rule using our grid

```
# Compute normalized marginalized posteriors, P(a|D) and P(b|D)
# by summing over other parameter. Normalize by gridding.
p_a_D <- apply(exp(z), 1, sum)
p_a_D <- p_a_D/(delta_a*sum(p_a_D))
p_b_D <- apply(exp(z), 2, sum)
p_b_D <- p_b_D/(delta_b*sum(p_b_D))

# Compute normalized conditional posteriors, P(a|b,D) and P(b|a,D)
# using true values of conditioned parameters. Vectorize(func, par)
# makes a vectorized function out of func in the parameter par.
p_a_bD <- exp(Vectorize(log.post, "a")(ddat, xdat, a, B.true,
                                        x0, w, Delta.t))
p_a_bD <- p_a_bD/(delta_a*sum(p_a_bD))
p_b_aD <- exp(Vectorize(log.post, "b")(ddat, xdat, A.true, b,
                                        x0, w, Delta.t))
p_b_aD <- p_b_aD/(delta_b*sum(p_b_aD))
```

Signal+Background - marginalization

```
par(mfrow=c(2,2), mgp=c(2,0.8,0), mar=c(3.5,3.5,1,1), oma=0.1*c(1,1,1,1))

# Plot the 1D marginalized posteriors
plot(b, p_b_D, xlab="background,␣B", yaxs="i",
     ylim=1.05*c(0,max(p_b_D, p_b_aD)), ylab="P(B␣|␣D)␣␣and␣␣P(B␣|␣A,D)",
     type="l", lwd=2)
lines(b, p_b_aD, lwd=2, lty=2)
abline(v=B.true, col="grey")

plot(a, p_a_D, xlab="amplitude,␣A", yaxs="i",
     ylim=1.05*c(0,max(p_a_D, p_a_bD)), ylab="P(A␣|␣D)␣␣and␣␣P(A␣|␣B,D)",
     type="l", lwd=2)
lines(a, p_a_bD, lwd=2, lty=2)
abline(v=A.true, col="grey")
```



Best estimates and reliability

- finally we want to get the best estimate of our parameters from the inferred posterior

$$\mu_A = \int A \cdot P\left(A \mid D\right) dA \sim \Delta A \sum_{k=1}^{K} A \cdot P\left(A_j \mid D\right)$$

$$\sigma_A^2 = \int (A - \mu_A)^2 \, P\left(A \mid D\right) dA \sim \Delta A \sum_{k=1}^{K} (A_k - \mu_A)^2 \, P\left(A_j \mid D\right)$$

- and

$$\begin{aligned} \mathrm{Cov}(A,B) &= \iint (A - \mu_A)\,(B - \mu_B)\, P\left(A, B \mid D\right) dA dB \\ &\sim \sum_{j=1}^{K} \sum_{k=1}^{K} \left(A_j - \mu_A\right)(B_k - \mu_B)\, P\left(A_j, B_k \mid D\right) \\ \rho &= \frac{\mathrm{Cov}(A,B)}{\sigma_A \sigma_B} \end{aligned}$$

- **Note**: there should be a factor $K/(K-1)$ in the variance and covariance calculations, but we can neglect it due to the large number of points in the grid (K)

Signal+Background - marginalization

```
# Compute normalized marginalized posteriors, P(a|D) and P(b|D)
# by summing over other parameter. Normalize by gridding.
p_a_D <- apply(exp(z), 1, sum)
p_a_D <- p_a_D/(delta_a*sum(p_a_D))
p_b_D <- apply(exp(z), 2, sum)
p_b_D <- p_b_D/(delta_b*sum(p_b_D))

# Compute mean, standard deviation, covariance, correlation, of A and B
mean_a <- delta_a * sum(a * p_a_D)
mean_b <- delta_b * sum(b * p_b_D)
sd_a   <- sqrt( delta_a * sum((a-mean_a)^2 * p_a_D) )
sd_b   <- sqrt( delta_b * sum((b-mean_b)^2 * p_b_D) )

# Covariance nurmalization is performed with 'brute force'
# The normalization constant is Z = delta_a*delta_b*sum(exp(z)).
# This is independent of (a,b) so can be calculated outside of the loops.
cov_ab <- 0
for(j in 1:length(a)) {
    for(k in 1:length(b)) {
        cov_ab <- cov_ab + (a[j]-mean_a)*(b[k]-mean_b)*exp(z[j,k])
    }
}
cov_ab <- cov_ab / sum(exp(z))
rho_ab <- cov_ab / (sd_a * sd_b)

cat("␣␣a␣=␣", mean_a, "+/-", sd_a, "\n")
cat("␣␣b␣=␣", mean_b, "+/-", sd_b, "\n")
cat("rho␣=␣", rho_ab, "\n")
```

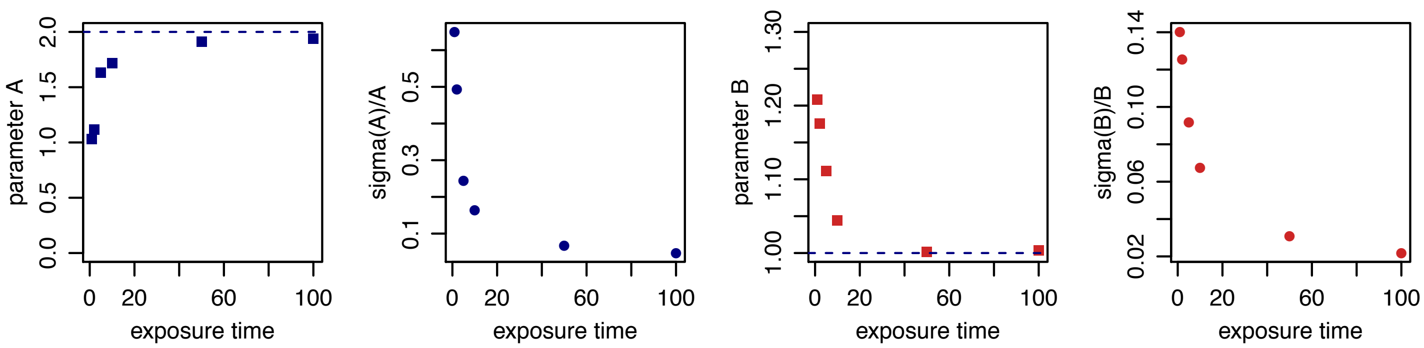
```
a = 1.630189 +/- 0.3983222
b = 1.111212 +/- 0.1020915
rho = -0.3968818
```


Signal+Background - further studies

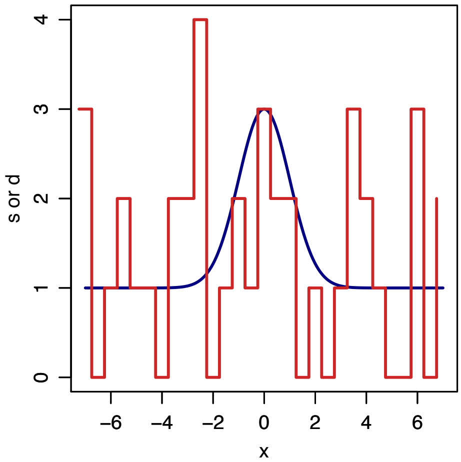
Dependence on exposure

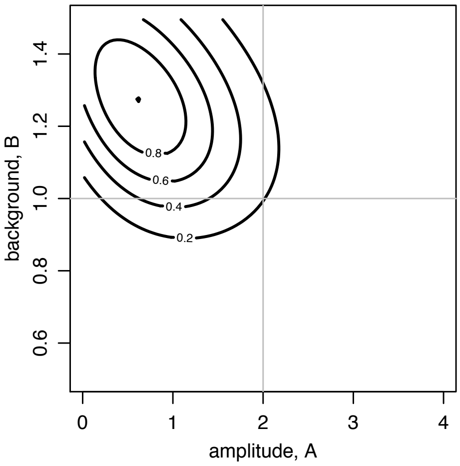
- the exposure time (Δt parameter) is directly connect to the number of collected photons
- larger exposures ➜ more collected photons
- smaller exposures ➜ decrease in accuracy and precision

Q: change the exposure to $\Delta t = \{1, 2, 10, 50, 100\}$ and check the effect on the results



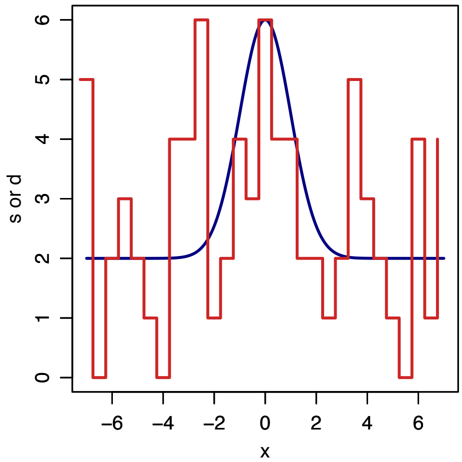
Signal+Background vs Exposure Time

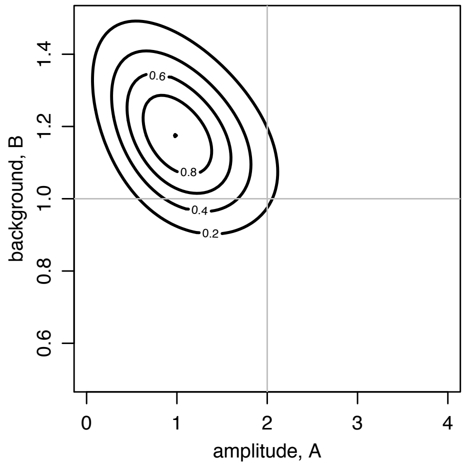




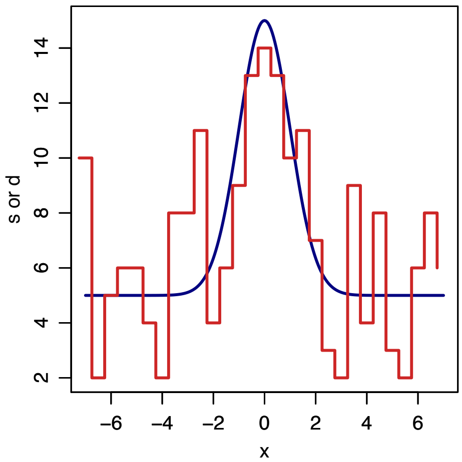
$\Delta t = 1$ a.u.

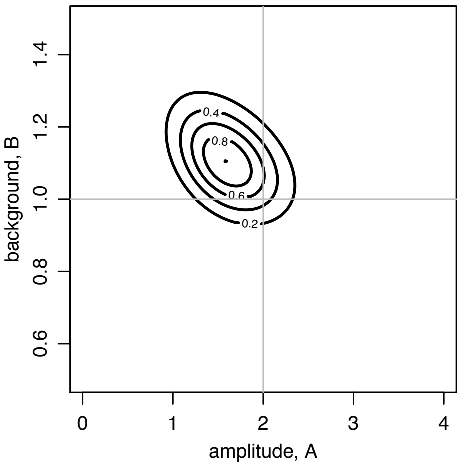
$\Delta t = 2$ a.u.





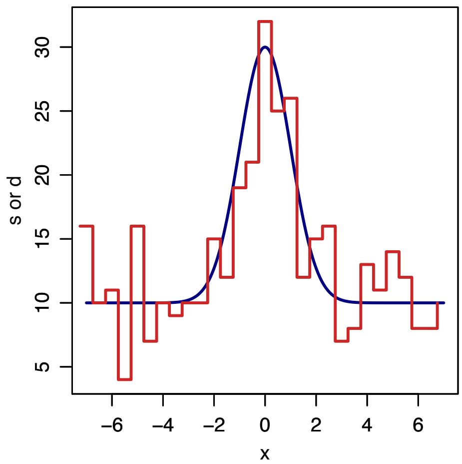
Signal+Background vs Exposure Time

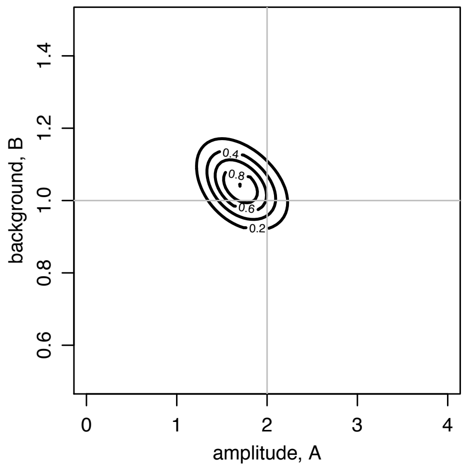




$\Delta t = 5$ a.u.

$\Delta t = 10$ a.u.





Signal+Background : other possible studies

Dependence on resolution

- vary the sampling resolution of used to generate the data, keeping the same sampling range

  ```
  xdat <- seq(from=-7*w, to=7*w, by=0.5*w)
  ```

- change the resolution $w = \{0.1, 0.25, 1, 2, 3\}$

Q: Check the effect on the results

Dependence on resolution

- change the ratio A/B used to simulate the data (keeping both positive in accordance with the prior)

Q: Check the effect on the results