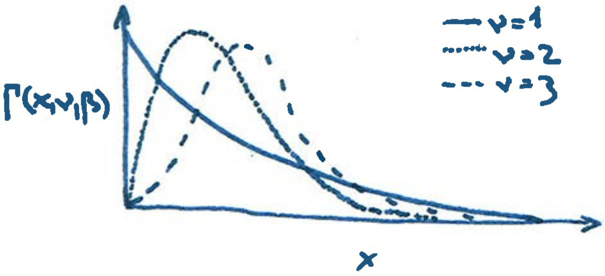
The Gamma distribution

- The Gamma distribution of a random variable is

$$\mathcal{P}(x|\nu,\beta) = \frac{1}{\Gamma(\nu)} \beta^{-\nu} x^{\nu-1} \exp(-\frac{x}{\beta})$$

- Sketch it for $\nu = 1,2,3$



$\nu=1$: $x^{1-1} = x^0 \Rightarrow \exp(-\frac{x}{\beta})$ survives

$\nu=2$: $x^{2-1} = x \Rightarrow x\exp(-\frac{x}{\beta})$

$\nu=3$, $x^2\exp(-x/\beta)$

- For a single draw, which is the most likely value of the random variable?

$\rightsquigarrow$ x is the random variable

$\rightsquigarrow$ „most likely": search maximum

$\rightsquigarrow$ $\frac{\partial \mathcal{P}}{\partial x} \overset{!}{=} 0$ // search maximum

$\rightsquigarrow$ $\exp(\cdot)\left[(\nu-1)x^{\nu-2} + x^{\nu-1}(-\frac{1}{\beta})\right] \overset{!}{=} 0$

$\Leftrightarrow$ $\frac{(\nu-1)}{x} - \frac{1}{\beta} = 0$

$\Rightarrow$ $\underline{x = \beta(\nu-1)}$

- The most likely value to be drawn is $x=\beta(\nu-1)$, which is zero for $\nu=1$, and >0 for $\nu>1$

- Someone gave you two measurements x_1 and x_2, and you know these were both independent draws from a Gamma-distribution, with the same ν, β.
- What is the joint distribution of x_1 and x_2?

$$\rightsquigarrow \, g(x_1, x_2) = g(x_1)\, g(x_2)$$

$$= \Gamma(x_1 | \nu, \beta) \cdot \Gamma(x_2 | \nu, \beta)$$

$$= \left[\frac{1}{\Gamma(\nu)} \beta^{-\nu}\right]^2 x_1^{\nu-1} x_2^{\nu-1} \exp\left(-\frac{x_1}{\beta}\right) \exp\left(-\frac{x_2}{\beta}\right)$$

$$= \left[\frac{1}{\Gamma(\nu)} \beta^{-\nu}\right]^2 (x_1 x_2)^{\nu-1} \exp\left(-\frac{1}{\beta}[x_1+x_2]\right)$$

- Given x_1 and x_2, what is the most likely value of β to have produced these measurement? If needed adopt flat priors.

$$\rightsquigarrow \, g(x_1, x_2 | \beta) = \left[\frac{1}{\Gamma(\nu)} \beta^{-\nu}\right]^2 (x_1 x_2)^{\nu-1} \exp\left(-\frac{1}{\beta}[x_1+x_2]\right)$$

$$\rightsquigarrow \, g(\beta | x_1, x_2) = \frac{g(x_1, x_2 | \beta)\, \pi(\beta)}{\pi(x_1, x_2)} \overset{\text{flat priors}}{\propto} \left[\frac{1}{\Gamma(\nu)} \beta^{-\nu}\right]^2 (x_1 x_2)^{\nu-1} \exp\left(-\frac{1}{\beta}[x_1+x_2]\right)$$

$\rightsquigarrow$ most likely value of β to have produced the measurements x_1 and x_2:

$$\frac{\partial}{\partial\beta} g(\beta | x_1, x_2) \overset{!}{=} 0$$

$$= \frac{(x_1 x_2)^{\nu-1}}{[\Gamma(\nu)]^2} \exp(\cdot) \left[(-2\nu)\beta^{-2\nu-1} + \beta^{-2\nu}[x_1+x_2]\beta^{-2}\right]$$

$$\Rightarrow (-2\nu)\beta^{-2\nu-1} + \frac{1}{\beta}\beta^{-2\nu-2}[x_1+x_2] \overset{!}{=} 0$$

$$\Rightarrow (-2\nu) + \frac{(x_1+x_2)}{\beta} \overset{!}{=} 0$$

$\Rightarrow$ The most likely value to have produced x_1 and x_2 was

$$\boxed{\beta = \frac{x_1+x_2}{2\nu} \quad \text{\% priors}}$$

- ## A simple derivation including a marginal

Ratio distributions

Someone gives you acess to two measurements, which produces values u and v. Your physical theory can only explain $y = u/v$. The two measurements u and v were independent and their uncertainties determine the (well known) distributions $\rho(u)$ and $\rho(v)$.

I made a mistake on the blackboard, where it said $y=uv$. The correct version is $y = u/v$

Derive the distribution $\rho(y)$.

$\Rightarrow \rho(u,v) = \rho(u)\,\rho(v)$; $y = u/v$, $v \neq 0$, so we can set [assuming everything positive]

$$\int \rho(y)\,dy = 1 = \int \rho(u,v)\,du\,dv$$

$$= \int \rho(yv, v)\,d(yv)\,dv$$

$$= \int \rho(yv,v) \left| \frac{d(yv)}{dy} \right| dy\,dv$$

$$= \int \rho(yv,v) \cdot v\,dy\,dv$$

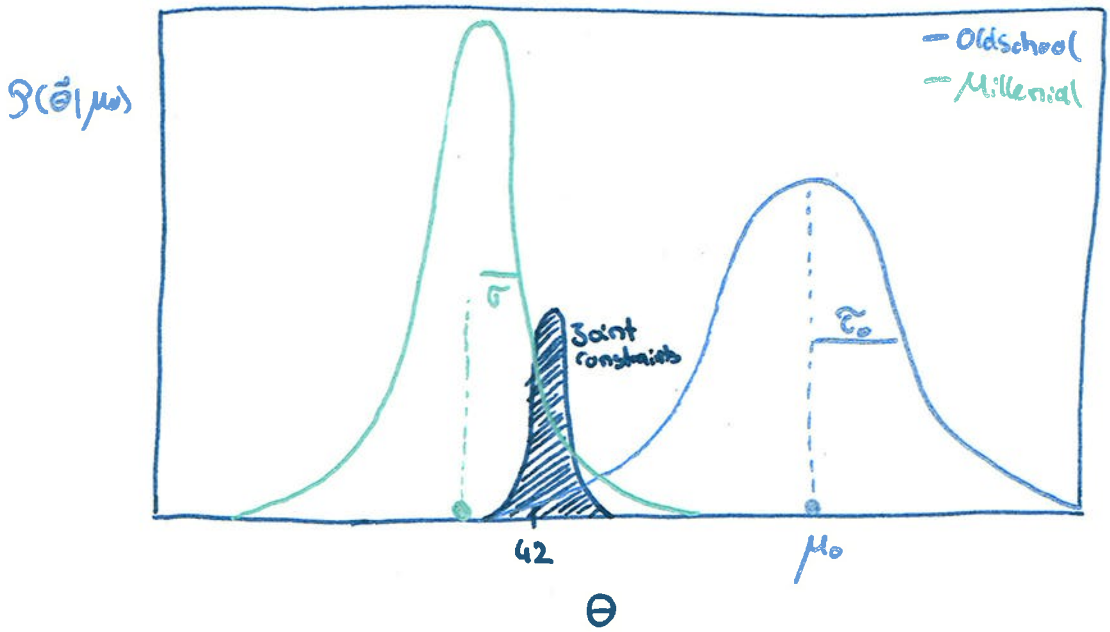
and hence by identification: $\cancel{\rho(y) = \int \rho(yv,v)}$

$$\rho(y) = \int \rho(yv,v) \cdot v\,dv$$

$$= \int \rho(yv)\,\rho(v)\,v\,dv \quad \text{for our indep variables}$$

Satellites: TNG

- Want: measure a parameter Θ
- Satellite „Oldschool" already put first constraints with uncertainty τ_0, and satellite „Millenials" shall do much better.



- Hyperparameters: parameters which are not of scientific interest, but which can be tuned (experimental design, see floor 11), or which were introduced to model „uninteresting physics" (then called nuisance parameters).

  ⇝ the hyperparams here are μ_0 and τ_0

- Derive the posterior $p(\Theta|y)$.

  Sampling distribution: $y \sim G(\Theta, \sigma^2)$, and we use the old posterior of „Oldschool" as a data-driven prior: $\pi(\Theta) = \cancel{G(\mu_0,\tau_0)}\ G(\Theta|\mu_0,\tau_0)$

So $p(y|\Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{[y-\Theta]^2}{\sigma^2}\right)$

and $\pi(\Theta) = \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{1}{2}\frac{[\Theta-\mu_0]^2}{\tau_0^2}\right)$

⇝ $p(\Theta|y) \propto p(y|\Theta)\,\pi(\Theta)$

$$p(\Theta|y) \propto \exp\left(-\frac{1}{2}\left[\frac{(y-\Theta)^2}{\sigma^2} + \frac{(\Theta-\mu_0)^2}{\tau_0^2}\right]\right)$$

$= G(\mu_1, \tau_1^2)$ with

$$\boxed{\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{y}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}} \qquad \boxed{\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}$$

Let's investigate what this means:

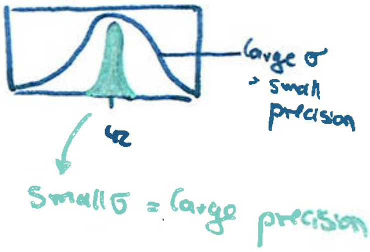
μ_1 : for $\sigma \ll \tau_0$: $\mu_1 \to \cancel{\mu_0}\ Y$ Millenials win

$\sigma \gg \tau_0$: $\mu_1 \to \mu_0$ oldschool wins

$\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$: σ large means $\frac{1}{\sigma^2}$ is small; σ small means $\frac{1}{\sigma^2}$ large $\Big\}$ $\frac{1}{\sigma^2}$ is called „precision"

„precisions of independent measurements are additive"

clear:

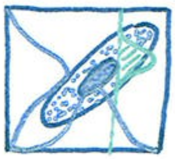


~> depending on σ and τ_0, the final estimate of y will be closer to the peak of Oldschool ($\tau_0 < \sigma$) or Millenial ($\sigma < \tau_0$).

~> since $\frac{1}{\tau_1^2} > \frac{1}{\tau_0^2}$ or $\frac{1}{\sigma^2}$, the combination of experiments achieves a tighter constraint

And! We saw why people tend to add variances together, or in this case inverse variances: This follows from Gaussians being self-conjugate, i.e. all our probability manipulations for Gaussians again produced Gaussians

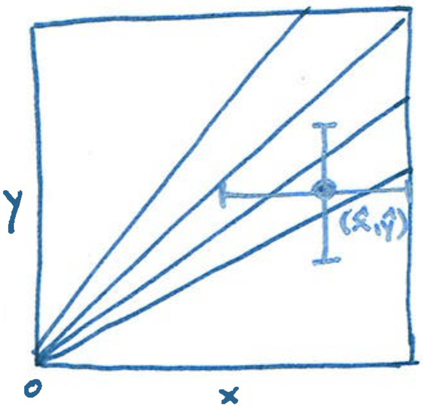
- Multivariate Gaussians have Gaussian conditionals and marginals
  ~> see lecture of the 14th Feb



- Products of Gaussians are again Gaussians
  ~> completion of the square

~> variances or precisions add (sometimes in weighted ways)

Straight line fitting with a twist



$y = mx$ (i.e. all lines go through the origin

Sadly our observations are $\hat{x}$ and $\hat{y}$, which scatter around the x and y which are related by m.

- $\mathcal{P}(\hat{x}|x) = G(x, \sigma^2)$
- $\mathcal{P}(\hat{y}|y) = G(y, \sigma^2)$
- Want: $\mathcal{P}(m|\hat{x},\hat{y})$ in the long-run, a few preparatory considerations beforehand.

1) since $\hat{x}$, and $\hat{y}$ independent, have $\mathcal{P}(\hat{x},\hat{y}|x,y) = \mathcal{P}(\hat{x}|x)\,\mathcal{P}(\hat{y}|y)$.

⇝ we will need this in the upcoming derivation.

2) „Where is my physics?" ⇝ $\mathcal{P}(m|x,y) = \delta_D(m - y/x)$

⇝ will need this to split joint distributions intelligently

⇝ also: $\mathcal{P}(y|x,m) = \delta_D(y - mx)$

3) Inverting joint events: $(\hat{x},\hat{y})$ is a joint event, composed of the outcome $\hat{x}$ and $\hat{y}$. With Bayes' theorem we invert:

$$\mathcal{P}(\hat{x},\hat{y}|m) = \frac{\mathcal{P}(m|\hat{x},\hat{y})\,\pi(\hat{x},\hat{y})}{\pi(m)}$$

⇝ with this preparation we can now work out $\mathcal{P}(m|\hat{x},\hat{y})$, of which there are many ways (see latex notes), the result being:

check: is it the probability of our interesting parameter given all data we have? YES. ✓

Sometimes one gets additional priors.

Are my remaining integrations the number of latent variables minus the number of δ_D-functions minus the number of integrals which I did analytically? YES. ✓

$$\mathcal{P}(m|\hat{x},\hat{y}) = \int \frac{G(\hat{x}|x)\,G(\hat{y}|mx)\,\pi(x)\,\pi(m)}{\pi(\hat{x},\hat{y})}\,dx$$

Do I know all occurring distributions? Does my experiment // i.e. the „exercise setup" specify them? YES. ✓

Do I have a prior for my physical parameters? YES. ✓

this is the evidence; it normalizes the posterior: it is therefore a constant for inference and can hence be in or outside.

FAQ:
- How do I know this is the result? „——"
- Does $\pi(\hat{x},\hat{y})$ need to be in or outside of the integral? „——"

⇝ All 4 checks passed ⇒ this is the result, modulo mistakes of logic