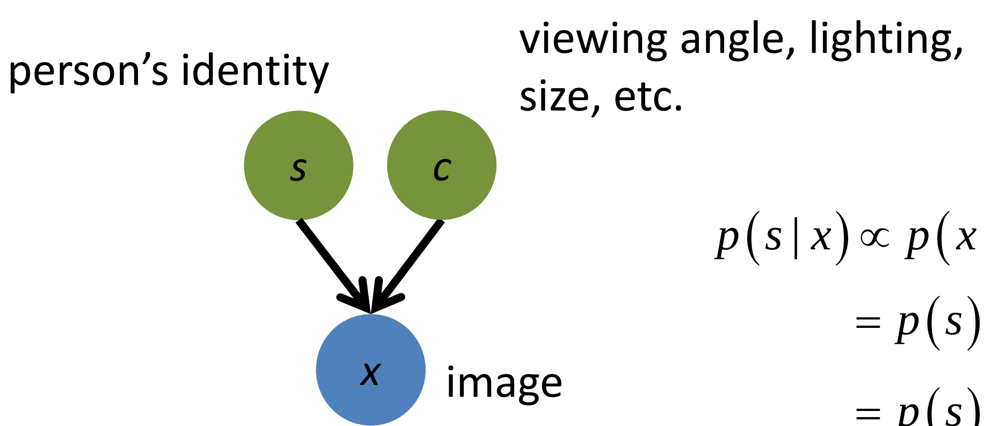
Homework set 2

(Lectures 3 and 4)

3.1 Recognizing a person

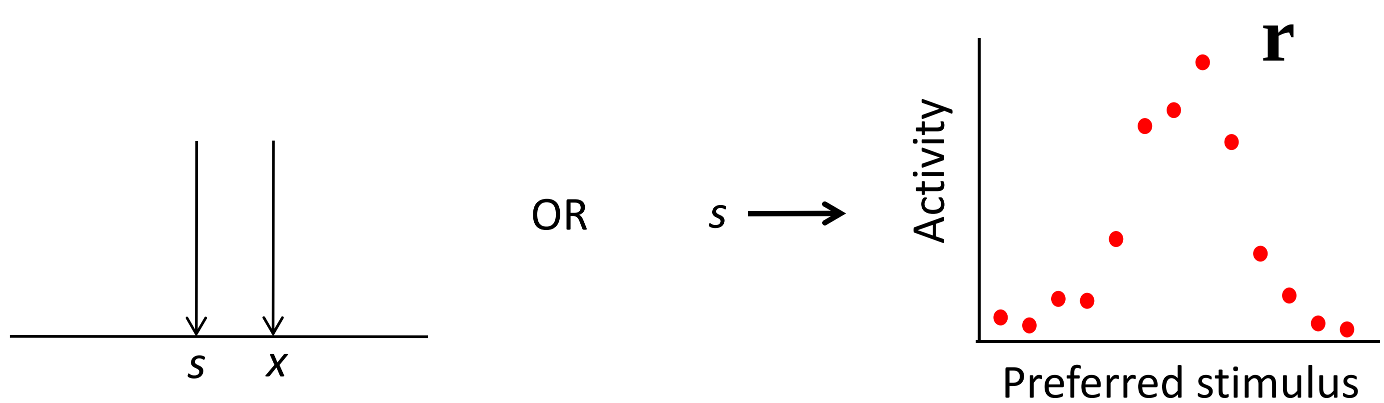
To recognize a person, we do not care about viewing angle and lighting conditions. Write down a generative model for this task and use it to explain how a Bayesian observer would infer a person's identity.

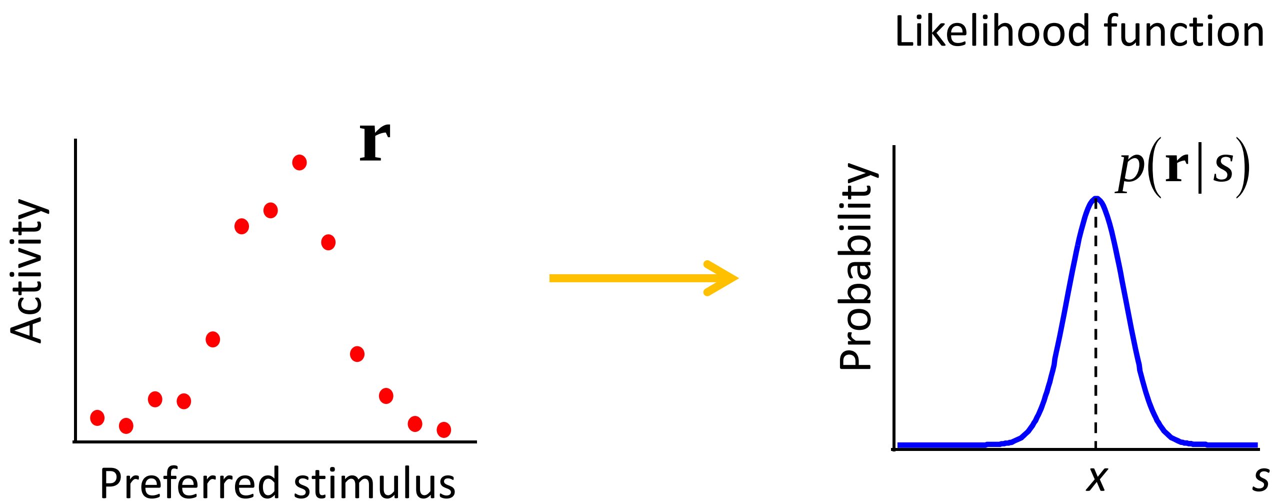


$$
\begin{aligned}
p(s \mid x) &\propto p(x \mid s)\, p(s) \\
&= p(s) \int p(x \mid s, c)\, p(c \mid s)\, dc \\
&= p(s) \int p(x \mid s, c)\, p(c)\, dc
\end{aligned}
$$

3.2 Internal representation vs. population activity

We have considered two different views of a perceptual observation: one is the "internal representation" of a stimulus. For example, if the true orientation of a line is 90°, then the internal representation might be 87°. The other view is the neural activity elicited by the stimulus, e.g. a population code in V1 representing the 90° line. How are these two views related? Is the population code more informative or not than the "internal representation"? Why?





$$x = \hat{s}_{\mathrm{ML}}\left(\mathbf{r}\right)$$

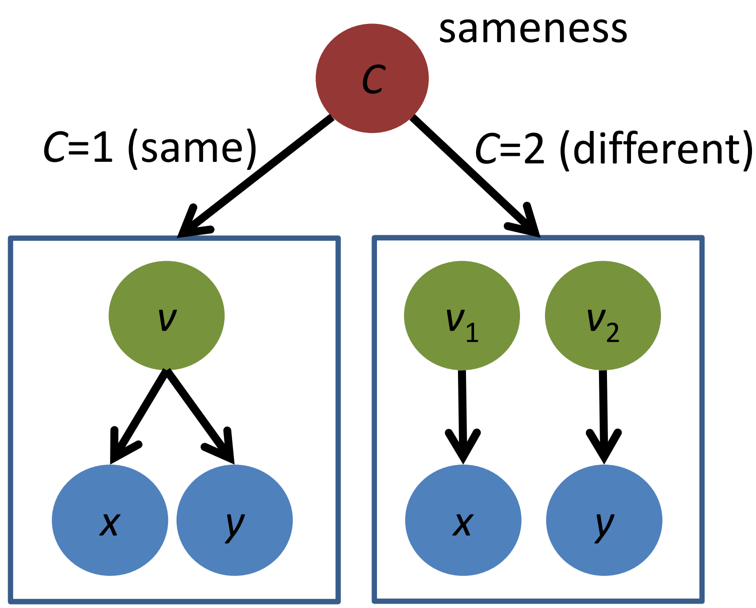
$$\mathbf{r} = \left(x, \sigma_{\mathrm{likelihood}}\right)$$

→ **r** is more informative than *x*

3.3 Same or different speeds?

Two objects are moving at unknown speeds. Denote speed by v and the noisy observations of the objects' speeds by x and y. Suppose the generative model, i.e. $p(x|v)$ and $p(y|v)$, as well as the prior distribution over speed, $p(v)$, are known. Using this knowledge, how does a Bayesian observer infer, based on x and y, whether the objects are moving at the same speed?

Bayesian model comparison (by the brain)



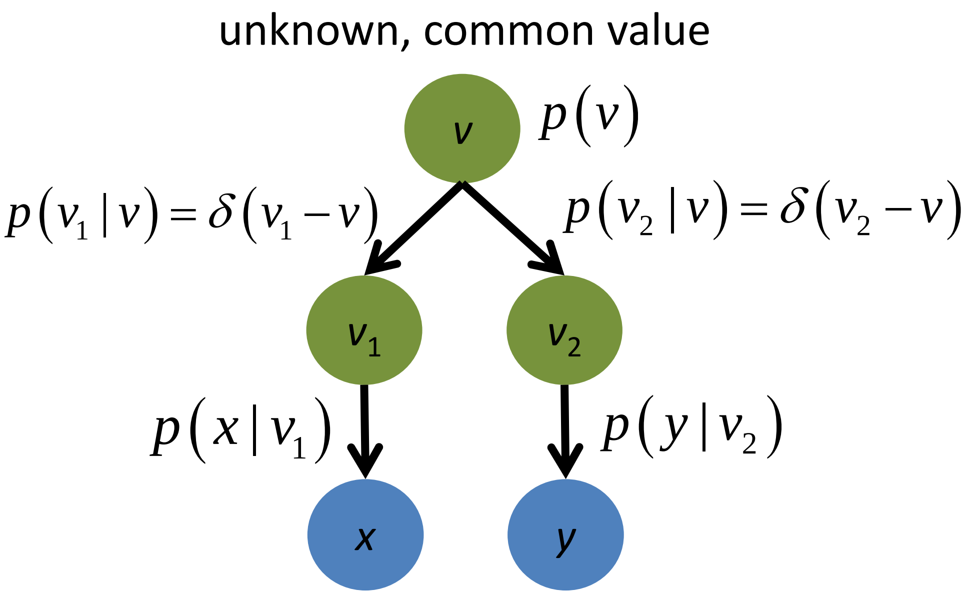
$$p(C\,|\,x,y) \propto p(x,y\,|\,C)\,p(C) = \frac{p(x,y\,|\,C)\,p(C)}{p(x,y\,|\,C=1)\,p(C=1) + p(x,y\,|\,C=2)\,p(C=2)}$$

$$p(x,y\,|\,C=1) = \int p(x,y\,|\,v)\,p(v)\,dv = \int p(x\,|\,v)\,p(y\,|\,v)\,p(v)\,dv$$

$$p(x,y\,|\,C=2) = \int\int p(x,y\,|\,v_1,v_2)\,p(v_1,v_2)\,dv_1dv_2 = \left(\int p(x\,|\,v_1)\,p(v_1)\,dv_1\right)\left(\int p(y\,|\,v_2)\,p(v_2)\,dv_2\right)$$

“But when the speeds are equal, the observations are still drawn from independent processes, that just happen to take the same value?”

Equal speed hypothesis

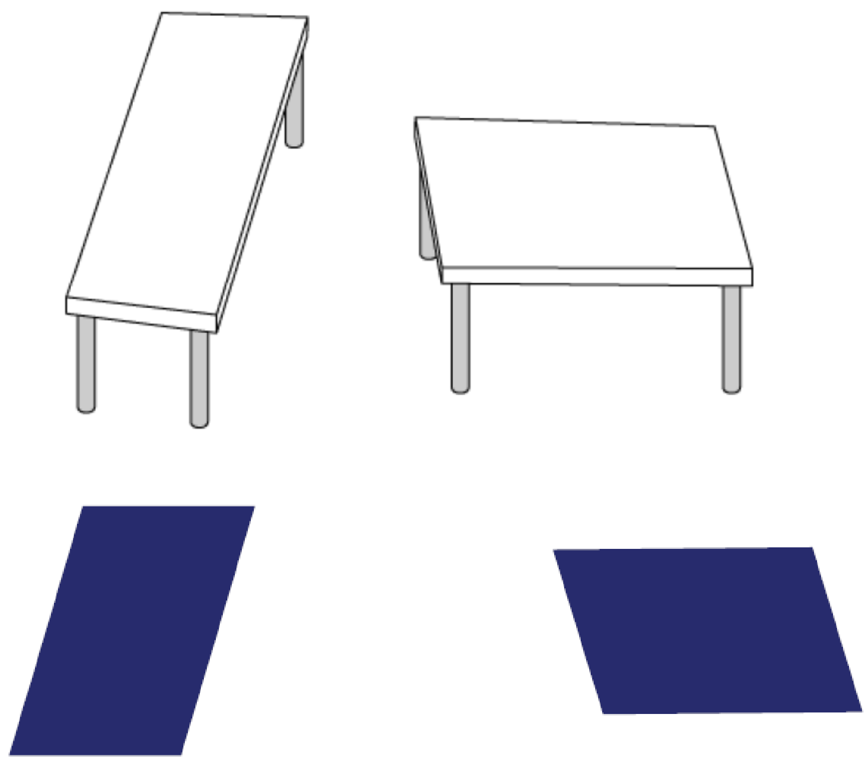


$$
\begin{aligned}
p(x, y | v_1 = v_2) &\propto \iiint p(x, y | v_1, v_2)\, p(v_1, v_2 | v)\, p(v)\, dv dv_1 dv_2 \\
&= \iiint p(x | v_1)\, p(y | v_2)\, \delta(v_1 - v)\, \delta(v_2 - v)\, p(v)\, dv dv_1 dv_2 \\
&= \int p(x | v_1 = v)\, p(y | v_2 = v)\, p(v)\, dv
\end{aligned}
$$

3.4 Visual illusions

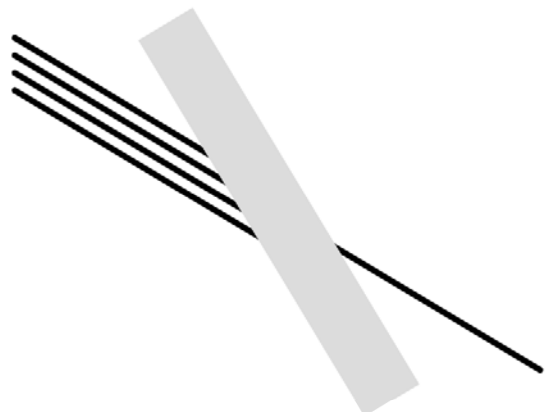
Find two visual illusions (other than the Ponzo and the horse/frog illusion, which were discussed in class). For each, draw a generative model to indicate the statistical dependencies between the variables. For each, use the generative model to write down a formal Bayesian model to explain the illusion. (The two illusions must have different generative models.)

Tabletop illusion



- Prior that parallellograms are rectangles lying flat
- Depth cue from angle

Poggendorf illusion



Overestimation of acute angles?

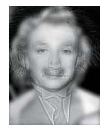
But…

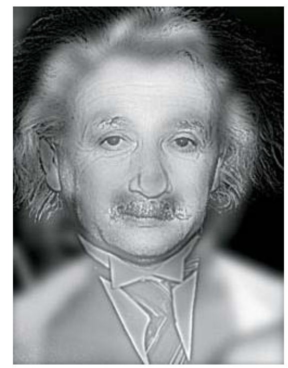
Hollow face illusion

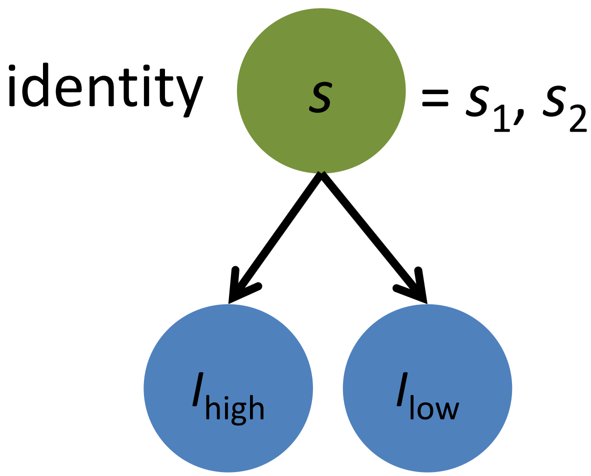
[YouTube video]

Convexity prior

Monroe vs Einstein



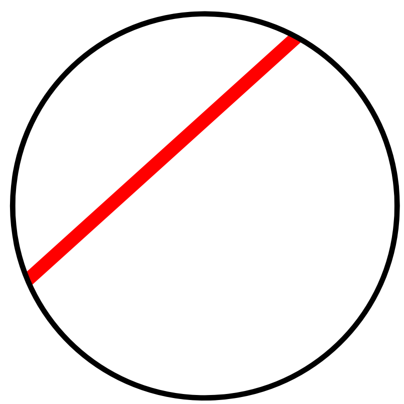




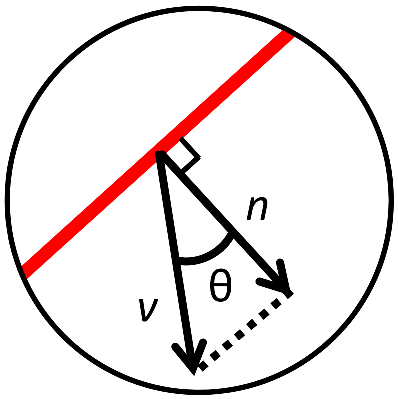
$$p\left(s_i \mid I_{\text{high}}, I_{\text{low}}\right) \propto p\left(I_{\text{high}} \mid s_i\right) p\left(I_{\text{low}} \mid s_i\right)$$

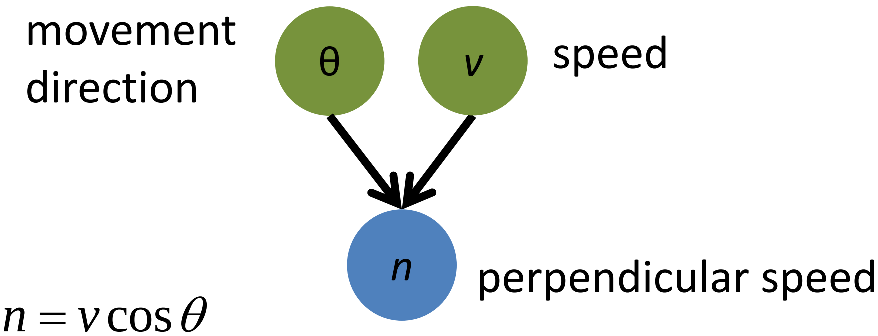
3.5 Aperture problem

A bar moving behind a circular aperture is generally perceived as moving in a direction perpendicular to its own orientation, even though its motion is consistent with a wide range of directions. Explain this in a Bayesian framework using a prior preference for low speeds.



Aperture problem





$$n = v\cos\theta$$

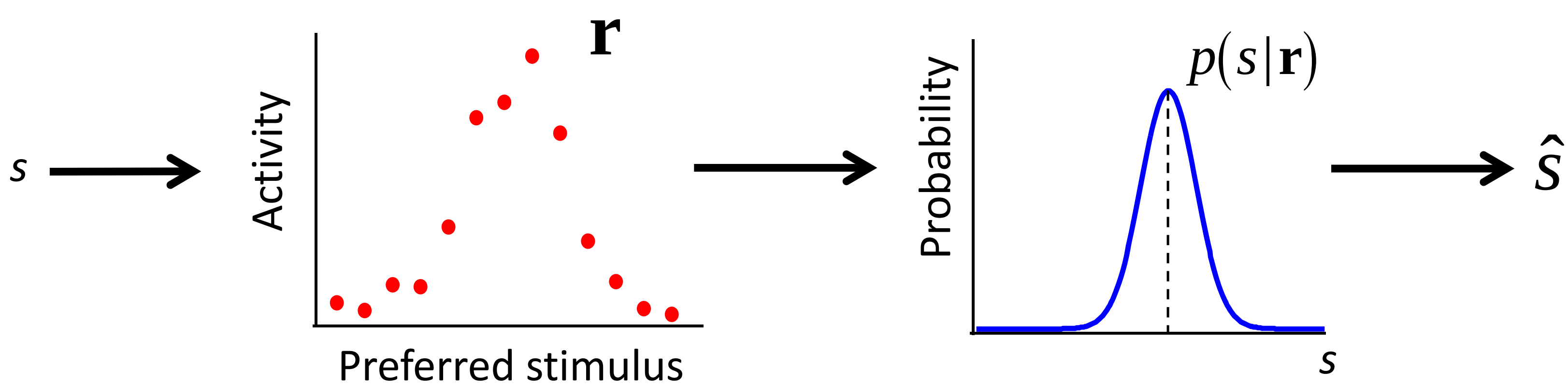
$$p(n \mid v,\theta) = \delta(n - v\cos\theta)$$

$$\begin{aligned} p(\theta \mid n) &\propto p(n \mid \theta) \\ &= \int p(n \mid v,\theta)\, p(v)\, dv \\ &= \int \delta(n - v\cos\theta)\, p(v)\, dv \\ &= \int \delta\left(v - \frac{n}{\cos\theta}\right) p(v)\, dv \\ &= p\left(v = \frac{n}{\cos\theta}\right) \end{aligned}$$

Largest for low speeds, i.e. when $\frac{1}{\cos\theta}$ is minimal.

4.1 Response (estimate) distribution

What is the general equation for the response distribution (assuming some deterministic decoder) in terms of the posterior distribution?



$$\hat{s} = \Phi\left(p\left(s \mid \mathbf{r}\right)\right) = F\left(\mathbf{r}\right)$$

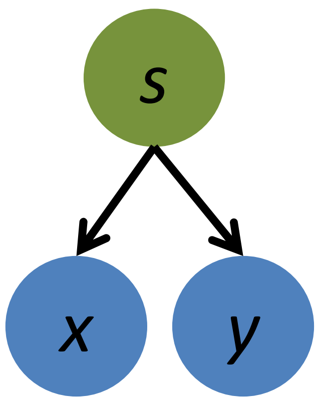
$$p\left(\hat{s} \mid s\right) = \int \delta\left(\hat{s} - F\left(\mathbf{r}\right)\right) p\left(\mathbf{r} \mid s\right) d\mathbf{r}$$

delta function in $\hat{s}$, but usually not in **r** !

4.2 Posterior vs response distribution

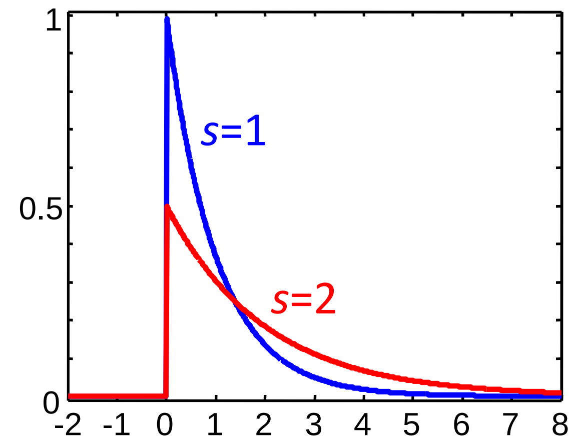
Work out a case where the posterior distribution and the response distribution are both continuous but very different from each other. (For example, choose non-Gaussian distributions and/or a more complex generative model.)

Gamma distributions

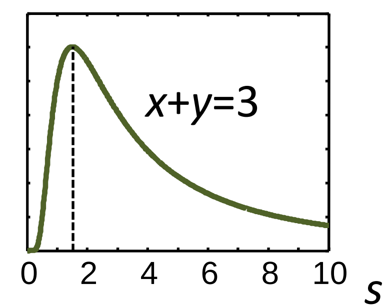


$$p(x\,|\,s)=\begin{cases}\frac{1}{s}e^{-\frac{x}{s}} & \text{for } x\geq 0\\ 0 & \text{for } x<0\end{cases}$$

$$p(y\,|\,s)=\begin{cases}\frac{1}{s}e^{-\frac{y}{s}} & \text{for } y\geq 0\\ 0 & \text{for } y<0\end{cases}$$

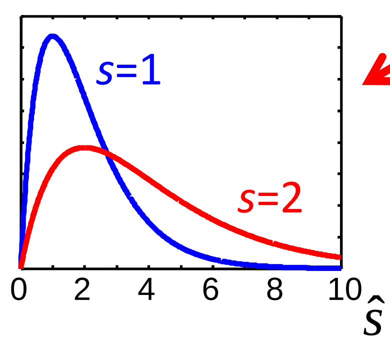


Posterior $$p(s\,|\,x,y)\propto p(x\,|\,s)\,p(y\,|\,s)=\frac{1}{s^2}e^{-\frac{x+y}{s}}$$



$$\hat{s}=\frac{x+y}{2}$$

Estimate distribution $$p(\hat{s}\,|\,s)\propto \hat{s}\,e^{-\frac{\hat{s}}{s}}$$



4.3 More posterior vs response distribution

Do the same when the stimulus variable is discrete, for example binary.

- Same as continuous, with discrete prior

$$p(s) = \frac{1}{2}\delta(s - s_1) + \frac{1}{2}\delta(s - s_2)$$

- Binary stimulus, binary observations

$$p(s = 1 \mid x_j, y_k) = \frac{p(x_j \mid s = 1)\, p(y_k \mid s = 1)}{p(x_j \mid s = 0)\, p(y_k \mid s = 0) + p(x_j \mid s = 1)\, p(y_k \mid s = 1)}$$

4.5 Simple inference with prior

Even when a single stimulus has to be inferred from a single cue, a bias can arise due to a prior. Assuming a Gaussian noise model and a Gaussian prior (with specified mean and variance), compute the bias as a function of the stimulus.

$$p(x|s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-s)^2}{2\sigma^2}} \qquad p(s) = \frac{1}{\sqrt{2\pi\sigma_p^2}} e^{-\frac{(s-\mu)^2}{2\sigma_p^2}}$$

$$p(s|x) \propto p(x|s)\, p(s) \propto e^{-\frac{(x-s)^2}{2\sigma^2}} e^{-\frac{(s-s_p)^2}{2\sigma_p^2}} \propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_p^2}+\frac{1}{\sigma^2}\right)\left(s-\frac{\sigma_p^2 x+\sigma^2\mu}{\sigma_p^2+\sigma^2}\right)^2\right]$$

Maximum at $\hat{s} = \frac{\sigma_p^2 x + \sigma^2 \mu}{\sigma_p^2+\sigma^2}$ Bias: $b(s) = \langle \hat{s} \rangle - s = \frac{\sigma_p^2 s + \sigma^2\mu}{\sigma_p^2+\sigma^2} - s = \frac{\sigma^2}{\sigma_p^2+\sigma^2}(\mu - s)$