2. Decision Theory

2.1 Optimal Risk
Rational decisions are informed by possible consequences of actions

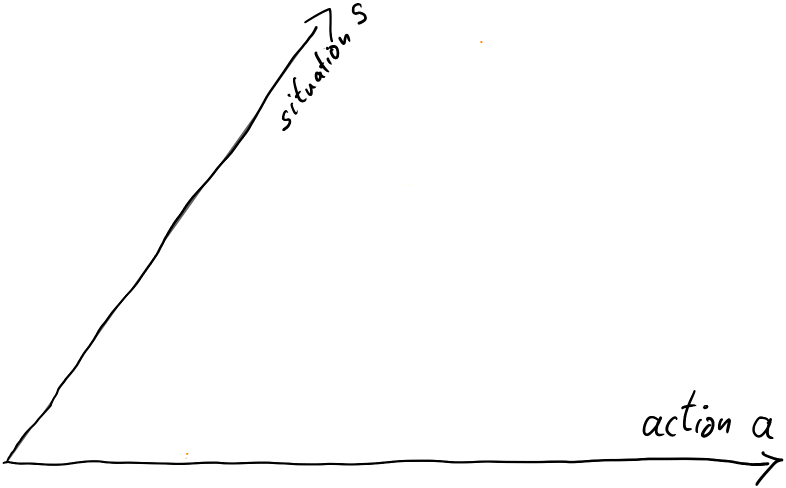
Loss function $l(a, s)$ quantifies loss of action a in situation s,
e.g. lost money, status, health, security, attention, or combination thereof
unfortunately: s is unknown

Risk of action a given data d on s is expected loss

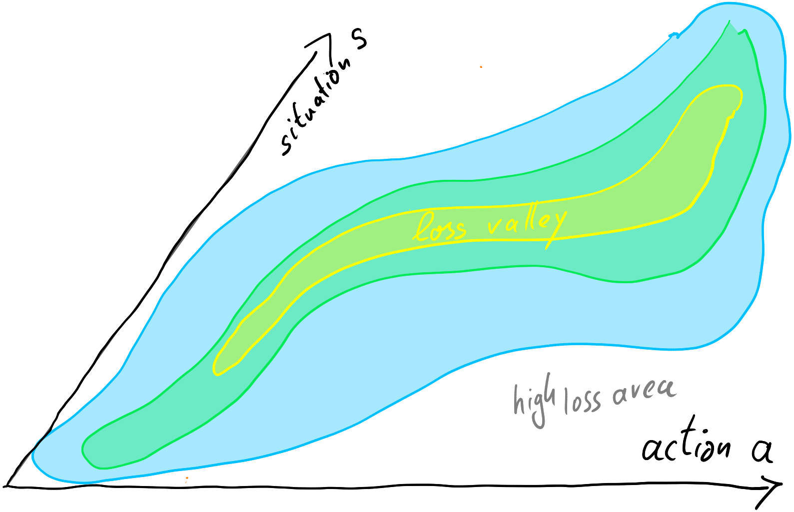
$$r(a,\, d) = \langle l(a,\, s) \rangle_{(s|d)} = \int ds\, l(a,\, s) \mathcal{P}(s|d)$$

Optimal risk: minimal risk provided by **optimal action**

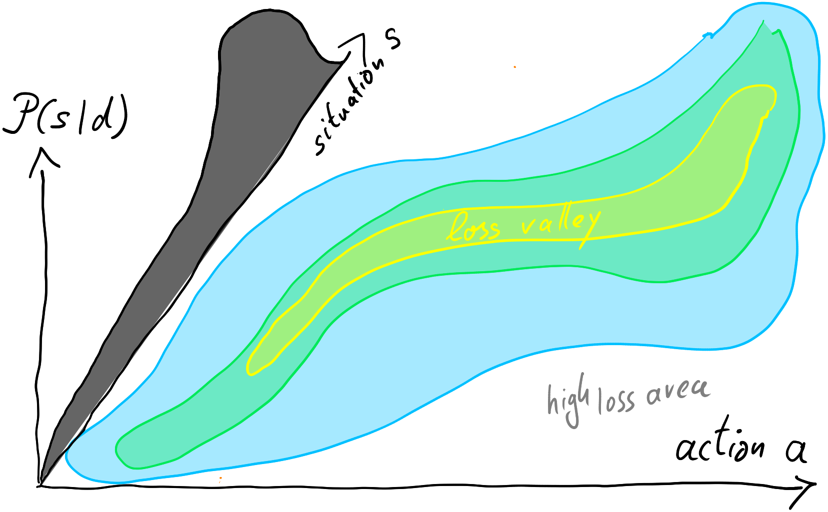
Optimal Risk



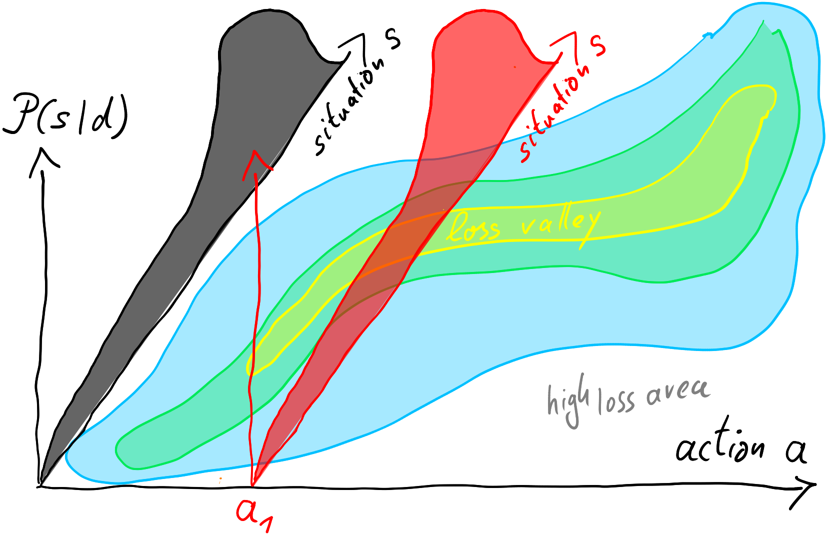
Optimal Risk



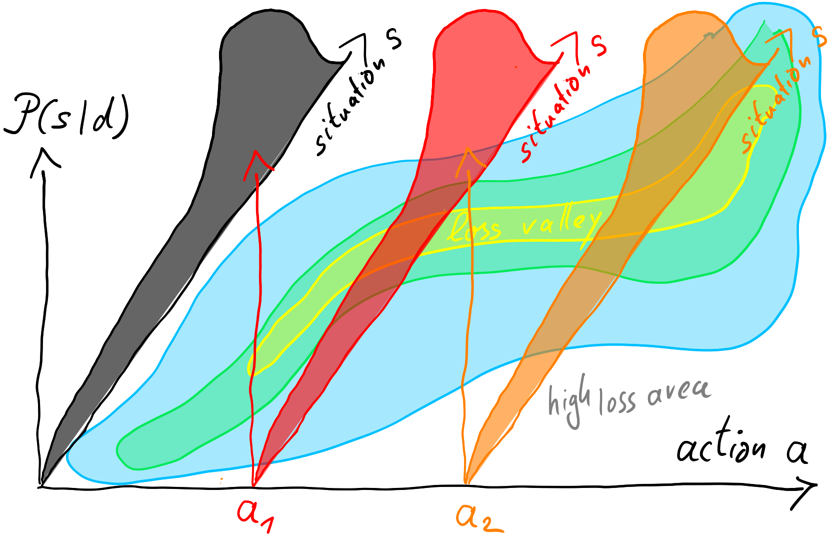
Optimal Risk



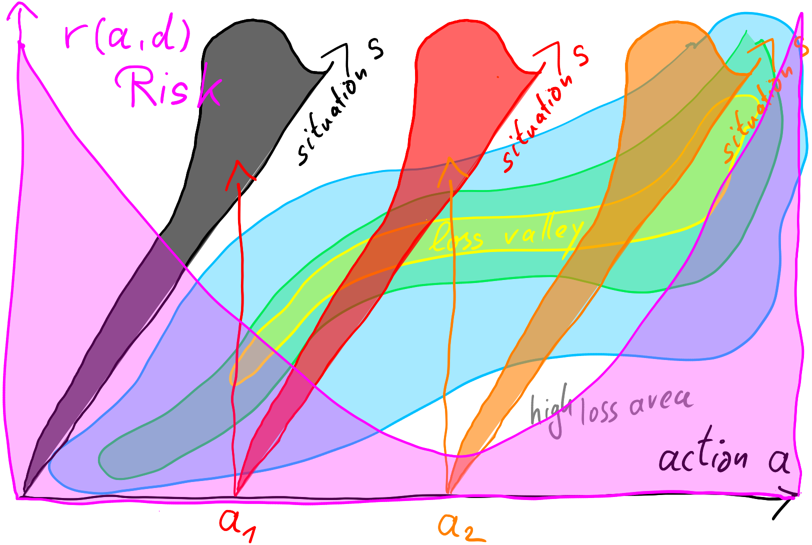
Optimal Risk



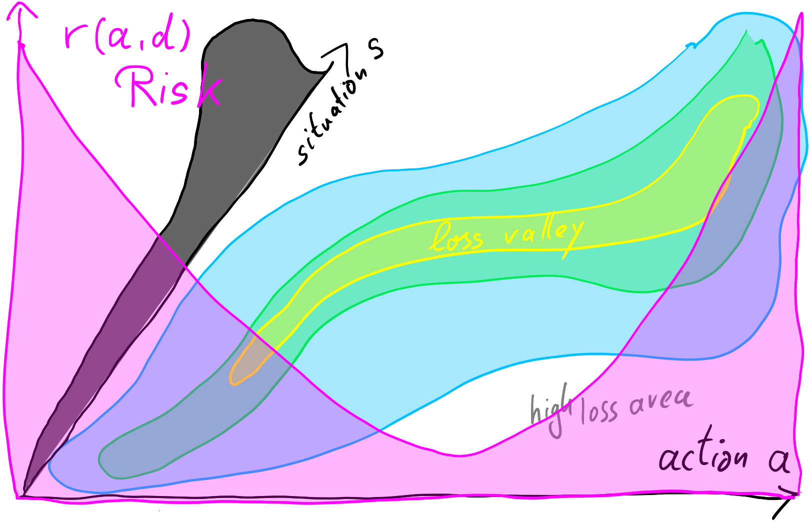
Optimal Risk



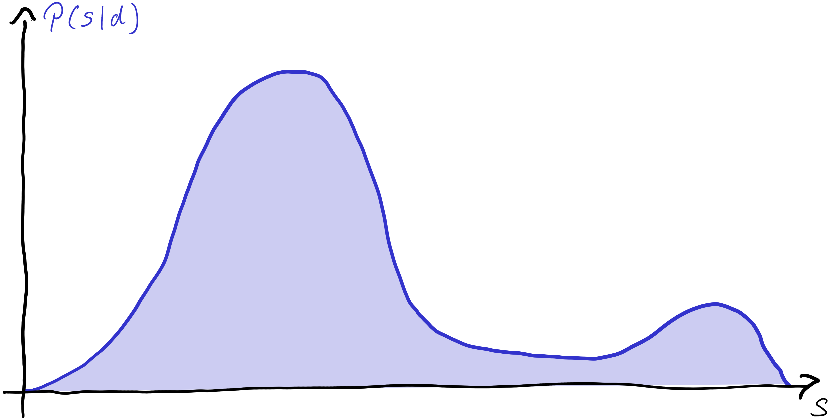
Optimal Risk



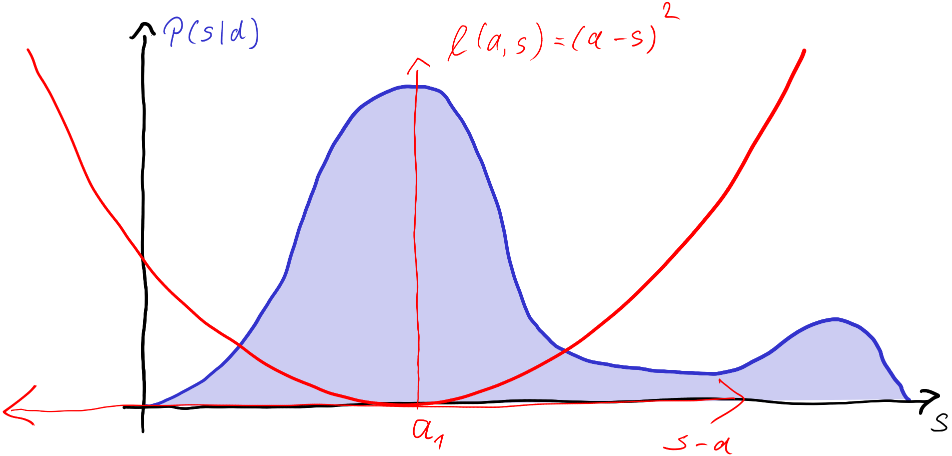
Optimal Risk



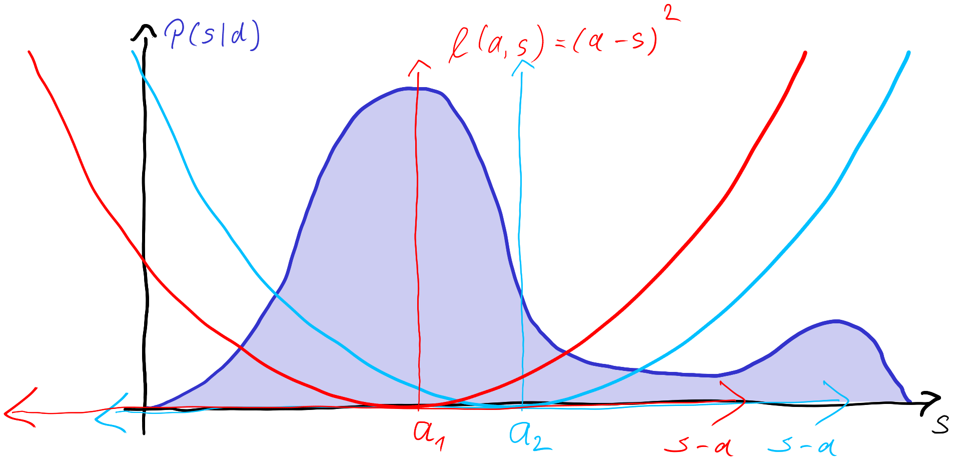
2.2 Loss Functions



Quadratic Loss



Quadratic Loss



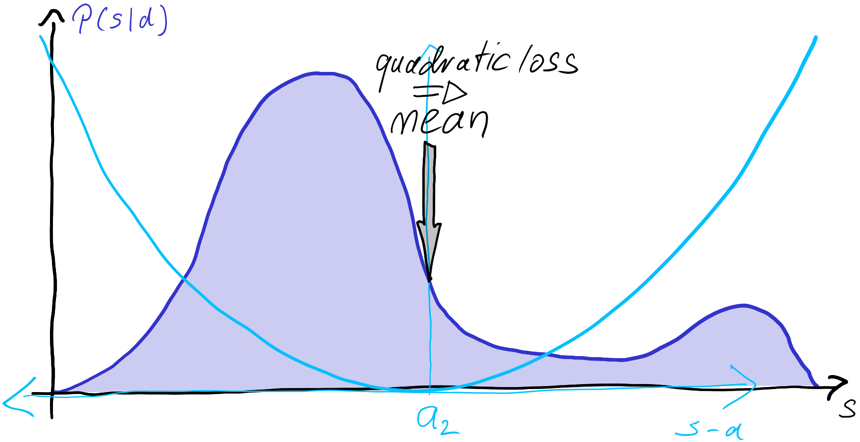
Quadratic Loss

used in science

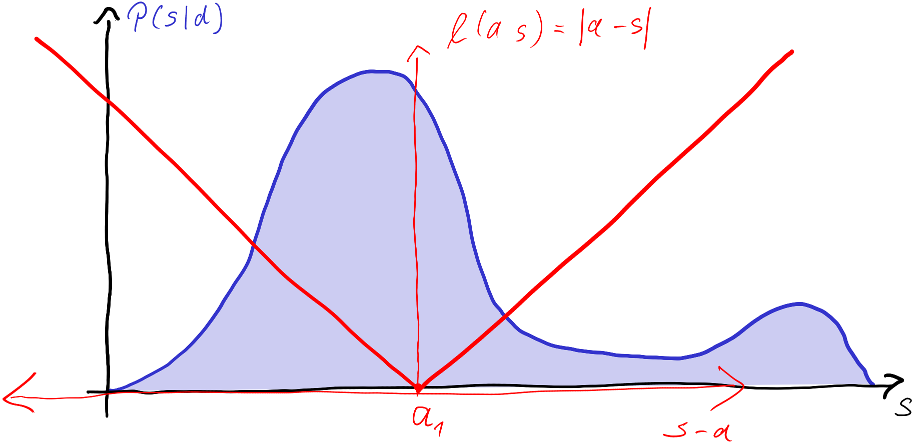
$$
\begin{aligned}
l(a,\, s) &= (a - s)^2 \\
r(a,\, d) &= \int ds\, (a - s)^2 \mathcal{P}(s|d) \\
&= \langle (a - s)^2 \rangle_{(s|d)} \\
\frac{\partial r(a,\, d)}{\partial a} &= \langle 2(a - s) \rangle_{(s|d)} \\
&= 2a - 2\langle s \rangle_{(s|d)} \stackrel{!}{=} 0 \\
\Rightarrow a &= \langle s \rangle_{(s|d)}
\end{aligned}
$$

Under quadratic loss best action is to choose **posterior mean**.

Quadratic Loss



Linear/Absolute Loss



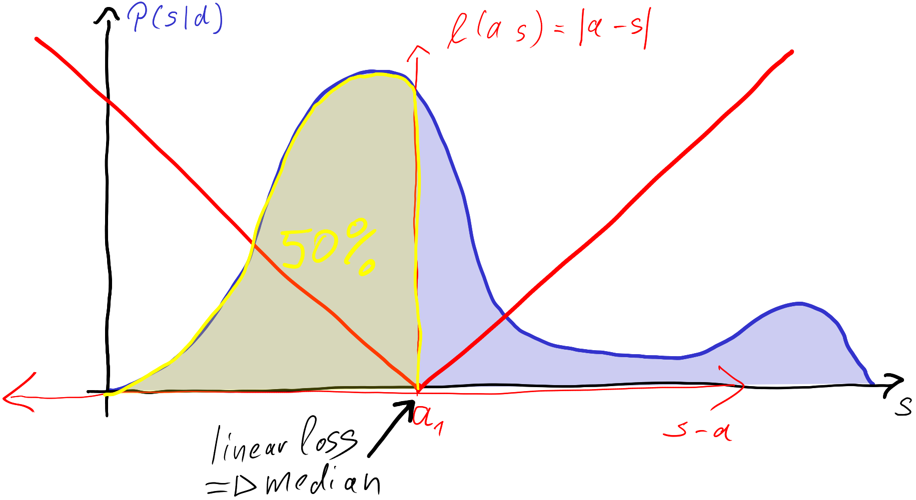
Linear/Absolute Loss

used in nummerics

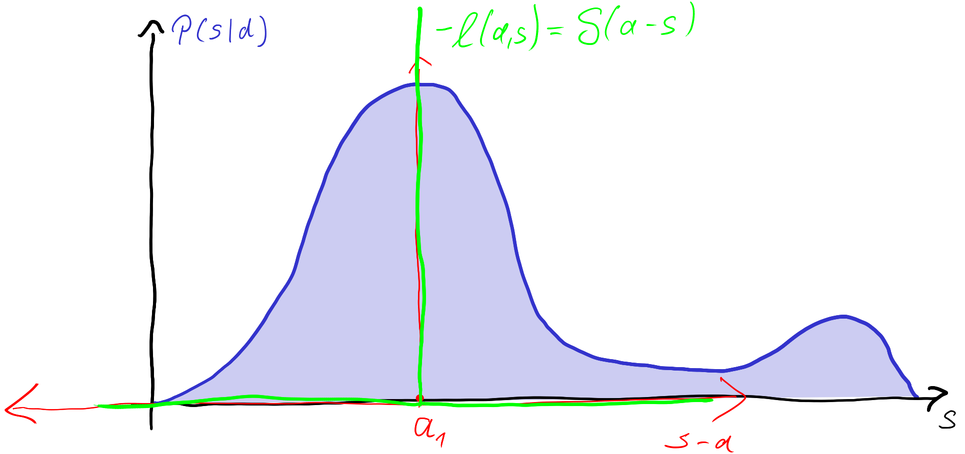
$$
\begin{aligned}
l(a, s) &= |a - s| \\
\frac{\partial r(a, d)}{\partial a} &= \frac{\partial}{\partial a} \int_{-\infty}^{\infty} ds\, |a - s|\, \mathcal{P}(s|d) \\
&= \frac{\partial}{\partial a} \left[\int_{-\infty}^{a} ds - \int_{a}^{\infty} ds \right] \left[(a - s) \mathcal{P}(s|d) \right] \\
&= \underbrace{\left[|_{s=a} + |_{s=a} \right] \left[(a - s) \mathcal{P}(s|d) \right]}_{=0} + \int_{-\infty}^{a} ds\, \mathcal{P}(s|d) - \int_{a}^{\infty} ds\, \mathcal{P}(s|d) \\
&\overset{!}{=} 0 \\
\Rightarrow \int_{-\infty}^{a} \mathcal{P}(s|d) &= \int_{a}^{\infty} \mathcal{P}(s|d) = 1/2
\end{aligned}
$$

Under linear loss the best action is to chose the **posterior median**.

Linear/Absolute Loss



Delta Loss



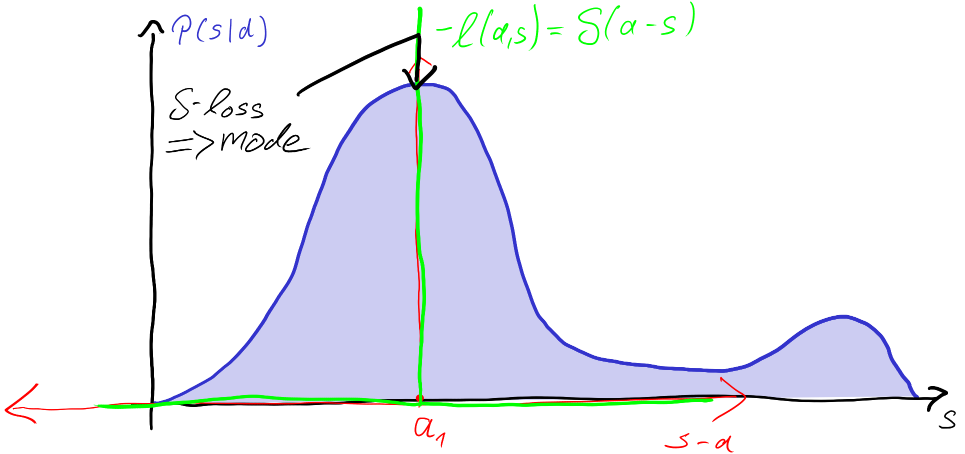
Delta Loss

used by military

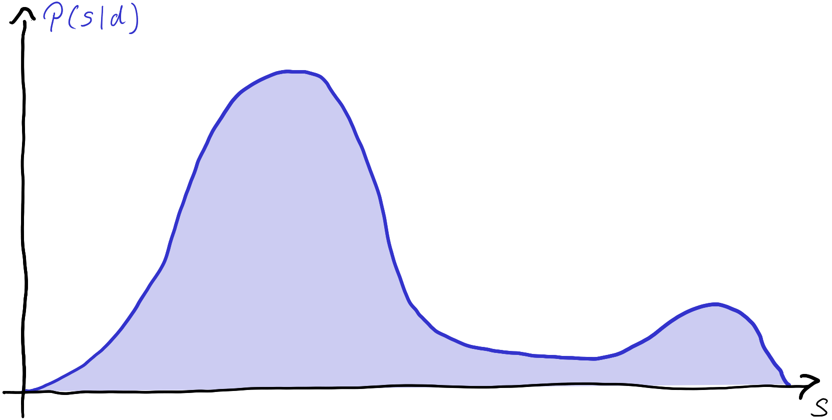
$$
\begin{aligned}
l(a,\, s) &= -\delta(a - s) \\
r(a, d) &= -\int ds\, \delta(a - s)\, \mathcal{P}(s|d) \\
&= -\mathcal{P}(a|d)
\end{aligned}
$$

Under delta loss the best action is to choose the **posterior mode**.

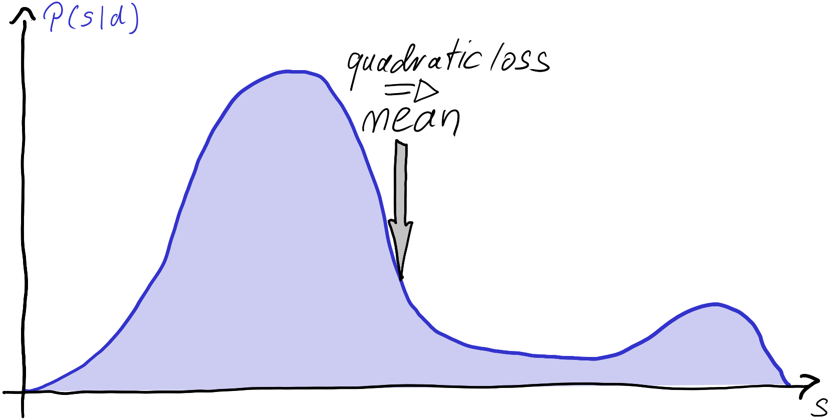
Delta Loss



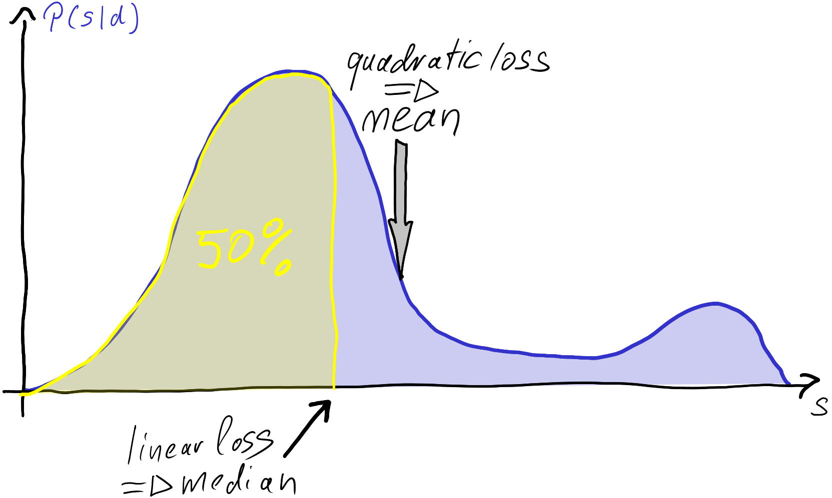
Mean, Median, & Mode



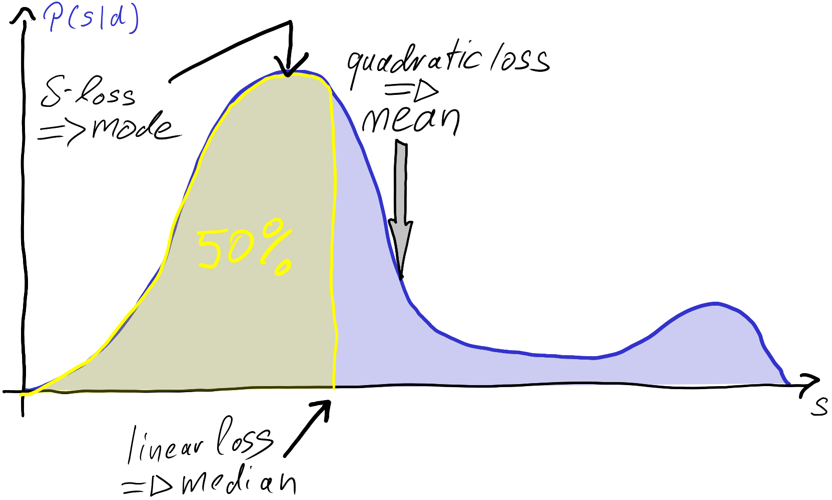
Mean, Median, & Mode



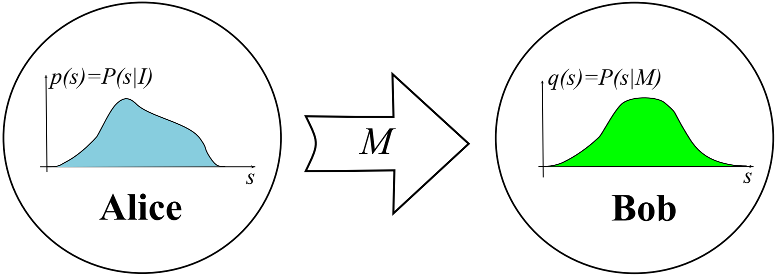
Mean, Median, & Mode



Mean, Median, & Mode



2.3 Communication



Alice: believe state $p(s) = \mathcal{P}(s|I)$ Bob: receiver state $q(s) = \mathcal{P}(s|M)$

Which message $M \in \mathbb{M}$ is to be sent?

Loss function needed: depends on cicumstances, in particular Bob's personal loss function

Alice: embarassement of being wrong, not to have assigned maximal probability to real s

2.2.1 Embarassement – a unique loss function

Leike & Enßlin (2017, Entropy, arXiv:161009018)

Alice chooses Bob's believe q through picking M. How to choose?

Wanted: Embarrassment function $l(q, s_0)$ for claiming $q(s)$ in case s turns out to be s_0.

$$\text{expected loss: } r(q, p) = \langle l(q, s_0) \rangle_{p(s_0)} = \int \mathrm{d}s_0 \, l(q, s_0) p(s_0)$$

Design criteria:

1. **Being local:** If $s = s_0$, only relevant prediction $q(s_0)$ matters, $l(q, s_0) = \mathscr{L}(q(s_0))$
   Counter example: $\widetilde{p}$-value $\widetilde{p}(s_0) = \int_{s \geq s_0} ds \, \mathcal{P}(s|I)$ depends on couterfactuals $s \neq s_0$
2. **Being proper:** If possible, $q = p$ should be communicated:
   $q = \mathrm{argmin_q} \, \langle \mathscr{L}(q(s_0)) \rangle_{p(s_0)} = p$

Normalization of q to be ensured, coordinate independence would be nice.

Determining Embarassement

$\mathscr{L}$ from minimizing $\langle \mathscr{L}(q(s_0)) \rangle_p$, ensuring normalization of q via Lagrange multiplier

$$
\begin{aligned}
0 &= \left(\underbrace{\frac{\mathrm{d}}{\mathrm{d}q(s)}}_{\text{minimum}} \left[\int \mathrm{d}s_0 \; \underbrace{\mathscr{L}(q(s_0))}_{\text{only } s_0 \text{ matters}} \; \underbrace{p(s_0)}_{\text{guessing } s_0} + \lambda \underbrace{\left(\int \mathrm{d}s_0 \, q(s_0) - 1 \right)}_{\text{normalization}} \right] \right)_{\underbrace{q=p}_{\text{properness}}} \\
&= \int \mathrm{d}s_0 \; \left[\mathscr{L}'(q(s_0)) \, \delta(s_0 - s) \, p(s_0) + \lambda \, \delta(s_0 - s) \right]_{q=p} \\
&= \left[\mathscr{L}'(p(s)) \, p(s) + \lambda \right]_{q=p} = \mathscr{L}'(p(s)) \, p(s) + \lambda \\
\Rightarrow \mathscr{L}'(x) &= -\frac{\lambda}{x} \Rightarrow \mathscr{L}(x) = -\lambda \ln(x) + \delta
\end{aligned}
$$

$\lambda > 0$, δ arbitrary constants *w.r.t.* q, *e.g.* $\lambda = 1$, $\delta = 0$

$$
r(q,p) = \langle l(q, s_0) \rangle_{p(s_0)} = \langle \mathscr{L}(q(s_0)) \rangle_{p(s_0)} = - \langle \ln(q(s_0)) \rangle_{p(s_0)}
$$

Embarassement & Entropy

$$\begin{aligned}
\textbf{Expected embarrassment: } r(p,q) &= \langle l(q,s_0)\rangle_{p(s_0)} \\
&= -\int \mathrm{d}s_0\, p(s_0)\ln(q(s_0)) \\
&= \mathcal{S}(p,q) \\
\textbf{Cross entropy: } \mathcal{S}(p,q) &:= -\int \mathrm{d}s\, p(s)\ln(q(s)) \\
\textbf{Entropy: } \quad \mathcal{S}(p) &:= -\int \mathrm{d}s\, p(s)\ln(p(s)) = \mathcal{S}(p,p)
\end{aligned}$$

Coordinate Invariance

$$
\begin{aligned}
\text{transformation } s \to s' &= s'(s),\ p(s) \to p'(s') = p(s)\left|\frac{ds}{ds'}\right|,\ q(s) \to q'(s') = q(s)\left|\frac{ds}{ds'}\right| \\
r(q',p') = \mathcal{S}(p',q') &= -\int \underbrace{ds'\, p'(s')}_{=ds\ p(s)} \ln q'(s') \\
&= -\int ds\, p(s)\left[\ln q(s) + \ln\left|\frac{ds}{ds'}\right|\right] \\
&= \mathcal{S}(p,q) - \underbrace{\left\langle \ln\left|\frac{ds}{ds'}\right| \right\rangle_{p(s)}}_{\text{const w.r.t. } q} \\
\text{in } l(q,s_0) &= \ln q(s_0) + \delta \text{ choose } \delta = -\mathcal{S}(p),\ \text{which is const w.r.t. } q \\
r(q,p) &= \mathcal{S}(p,q) - \mathcal{S}(p) = \int \mathrm{d}s\, p(s) \ln\left(\frac{p(s)}{q(s)}\right) \text{ is coordinate invariant}
\end{aligned}
$$

Kullback-Leiber Divergence

Kullback-Leibler divergence: $D_{\mathrm{KL}}(p||q) \;=\; \int \mathrm{d}s\, p(s) \ln\left(\frac{p(s)}{q(s)}\right) = \mathcal{S}(p,q) - \mathcal{S}(p)$

Gibbs inequality: $D_{\mathrm{KL}}(p||q) \geq 0$ if p and q are PDFs

$$\text{proof: } -D_{\mathrm{KL}}(p||q) = \int \mathrm{d}s\, p(s) \ln\left(\frac{q(s)}{p(s)}\right) \leq \int \mathrm{d}s\, p(s) \left(\frac{q(s)}{p(s)} - 1\right)$$
$$= \int \mathrm{d}s\, q(s) - \int \mathrm{d}s\, p(s) = 1 - 1 = 0 \quad \square$$

$D_{\mathrm{KL}}(p||q) = 0$ iff (if and only if) $q(s) = p(s)$ for $\forall s$
up to zero-measure differences (proof left for exercise)

End