Decision theory

Dr. Jarad Niemi

STAT 544 - Iowa State University

March 7, 2017

Bayesian statistician

Definition

A Bayesian statistician is an individual who makes decisions based on the probability distribution of those things we don't know conditional on what we know, i.e.

$$p(\theta|y, K).$$

Bayesian decision theory

Suppose we have an unknown quantity θ which we believe follows a probability distribution $p(\theta)$ and a decision (or action) δ. For each decision, we have a loss function $L(\theta, \delta)$ that describes how much we lose if θ is the truth. The expected loss is taken with respect to $\theta \sim p(\theta)$, i.e.

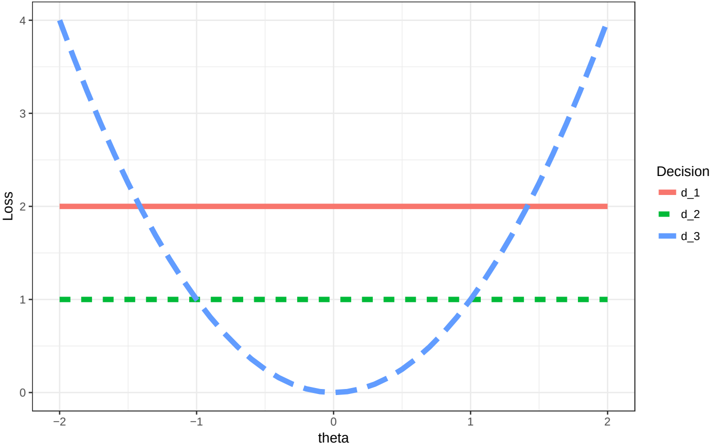
$$E_\theta[L(\theta, \delta)] = \int L(\theta, \delta)p(\theta)d\theta = f(\delta).$$

The optimal Bayesian decision is to choose δ that minimizes the expected loss, i.e.

$$\delta_{opt} = \text{argmin}_\delta E[L(\theta, \delta)] = \text{argmin}_\delta f(\delta).$$

Economists typically maximize expected utility where utility is the negative of loss, i.e. $U(\theta, \delta) = -L(\theta, \delta)$. If we have data, just replace the prior $p(\theta)$ with the posterior $p(\theta|y)$.

Depicting loss/utility functions



Parameter estimation

Definition

For a given loss function $L(\theta,\hat{\theta})$ where $\hat{\theta}$ is an estimator for θ, the Bayes estimator is the function $\hat{\theta}$ that minimizes the expected loss, i.e.

$$\hat{\theta} = \text{argmin}_{\hat{\theta}}\ E_{\theta|y}\left[L\left(\theta,\hat{\theta}\right)\middle| y\right].$$

Recall that

- $\hat{\theta} = E[\theta|y]$ minimizes $L(\theta,\hat{\theta}) = (\theta-\hat{\theta})^2$
- $0.5 = \int_{-\infty}^{\hat{\theta}} p(\theta|y)d\theta$ minimizes $L(\theta,\hat{\theta}) = |\theta-\hat{\theta}|$
- $\hat{\theta} = \text{argmax}_\theta p(\theta|y)$ is found as the minimizer of the sequence of loss functions $L(\theta,\hat{\theta}) = -\mathrm{I}(|\theta-\hat{\theta}|<\epsilon)$ as $\epsilon\to 0$

Which hand?

The setup:

- Randomly put a quarter in one of two hands with probability p.
- Let $\theta \in \{0,1\}$ indicate that the quarter is in the right hand.
- You get to choose whether the quarter is in the right hand or not.
- If you guess the quarter is in the right hand and it is, you get to keep the quarter. Otherwise, you don't get anything.

We have $\theta \sim Ber(p)$ and two actions

- a_0: say the quarter is not in the right hand and
- a_1: say the quarter is in the right hand.

Thus, the utility is

$$U(\theta, a_i) = \begin{cases} \$0.25\theta & \text{if } a_1 \\ 0 & \text{if } a_0 \end{cases}$$

and the expected utility is

$$E[U(\theta, a_i)] = \begin{cases} \$0.25p & \text{if } a_1 \\ 0 & \text{if } a_0 \end{cases}$$

So, we maximize expected utility by taking a_1 if $p > 0$.

How many quarters in the jar?

Suppose a jar is filled up to a pre-specified line. Let θ be the number of quarters in the jar. Provide a probability distribution for your uncertainty in θ. Suppose you choose

$$\theta \sim N(\mu, \sigma^2)$$

Since $\theta \in \mathbb{N}^+$, we can provide a formal prior by letting

$$P(\theta = q) \propto N(q; \mu, \sigma^2)\mathrm{I}(0 < q \le U)$$

for some upper bound U.

Guessing how many quarters are in the jar.

Now you are asked to guess how many quarters are in the jar. What should you guess?

Let q be the guess that the number of quarters is q, then our utility is

$$U(\theta, q) = q\mathrm{I}(\theta = q)$$

and our expected utility is

$$E_\theta[U(\theta, q)] = qP(\theta = q) \propto qN(q; \mu, \sigma^2)\mathrm{I}(0 \leq q \leq U).$$

Deriving the optimal decision

Here are three approaches for deriving the optimal decision:

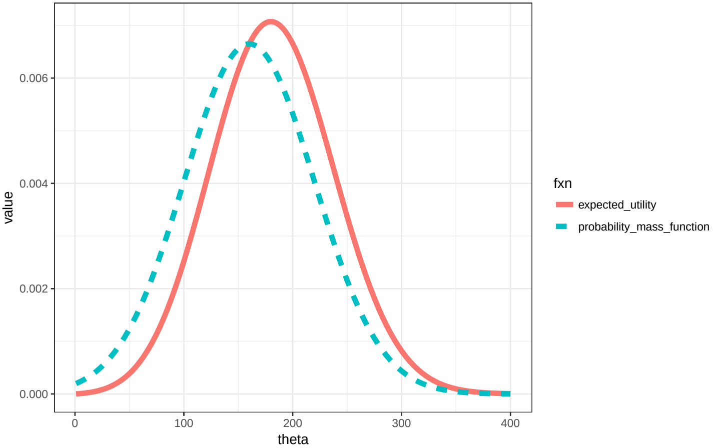
$$\text{argmax}_q f(q), \quad f(q) = qN(q;\mu,\sigma^2)\mathrm{I}(0 \le q \le U)$$

1. Evaluate $f(q)$ for $q \in \{1, 2, \ldots, U\}$ and find which one is the maximum.
2. Treat q as continuous and use a numerical optimization routine.
3. Take the derivative of $f(q)$, set it equal to zero, and solve for q.

In all cases, you are better off taking the $\log f(q)$ which is monotonic and therefore will still provide the same maximum as $f(q)$.

Visualizing the expected log utility

```
# p(theta) \propto N(theta;mu,sigma^2)I(1<= theta <= 400)
mu=160; sigma=60; U=400
```



Computational approaches

```r
log_f = Vectorize(function(q, mu, sigma, U) {
  if (q<0 | q>U) return(-Inf)
  return(log(q) + dnorm(q, mu, sigma, log=TRUE))
})

# Evaluate all options
log_expected_utility = log_f(1:U, mu=mu, sigma=sigma, U=U)
which.max(log_expected_utility) # since we are using integers 1:U
```

```
[1] 180
```

```r
# Numerical optimization
optimize(function(x) log_f(x, mu=mu, sigma=sigma, U=U), c(1,U), maximum=TRUE)
```

```
$maximum
[1] 180

$objective
[1] 0.1241182
```

Derivation

The function to maximize is

$$\log f(q) = \log(q) - (q-\mu)^2/2\sigma^2.$$

The derivative is

$$\frac{d}{dq}\log f(q) = \frac{1}{q} - (q-\mu)/\sigma^2.$$

Setting this equal to zero and multiplying by $-q\sigma^2$ results in

$$q^2 - \mu q - \sigma^2 = 0.$$

This is a quadratic with roots at

$$\frac{\mu \pm \sqrt{\mu^2 + 4\sigma^2}}{2}.$$

Since q must be positive, the answer is

```
(mu+sqrt(mu^2+4*sigma^2))/2
```

```
[1] 180
```

Sequential decisions

Consider a sequence of posteriors distributions $p(\theta_t|y_{1:t})$ that describe your uncertainty about the current state of the world θ_t given the data up to the current time $y_{1:t} = (y_1, \ldots, y_t)$. You also have a loss function for the current time $L(\theta_t, \delta_t)$. No suppose you are allowed to make a decision δ_{t+1} at each time t and this decision can affect the future states of the world θ_s for $s > t$.

At each time point, we have an optimal Bayes decision, i.e.

$$\text{argmin}_{\delta_{t+1}} \sum_{s=t+1}^{\infty} E_{\theta_s,\delta_s|y_{1:t}} \left[L\left(\theta_s, \delta_s\right) \middle| \, y_{1:t} \right].$$

But because your decision can affect future states which, in turn, can affect future decisions, your current decision needs to integrate over future decisions.