Lecture 2:
Bayesian inference - Discrete probability models

LINKÖPING UNIVERSITY

Many things about Bayesian inference for discrete probability models are similar to frequentist inference

Discrete probability models:

- Binomial sampling

Sampling a fix number of trials from a *Bernoulli process*

A Bernoulli process is a series of trials $(y_1, y_2, \ldots)$

- where in each trial
  - there two possible outcomes (*success* and *failure*)
  - the probability of success is constant = p
- where the members of the set of possible sequences $y_{(1)}, \ldots, y_{(M)}$ all with s successes and f failures ($s + f = M$) are ***exchangable***

The number of successes, $\tilde{r}$ in n trials is binomial distributed

$$P(\tilde{r} = r) = \binom{n}{r} p^r (1-p)^{n-r} = \frac{n!}{r!(n-r)!} \cdot p^r (1-p)^{n-r} \quad , r = 0,1,\ldots,n$$

- Hypergeometric sampling

Sampling a fix number n of items (without replacement) from a finite set of N items.

The finite set of items

- contains $Np = R$ items of a specific type ("success" item)

The number of success items, $\tilde{r}$ among the n sampled items is hypergeometric distributed

$$P(\tilde{r} = r) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}} \quad , r = 0,1,\ldots,\min(n, R)$$

- Pascal sampling

Sampling a random number of trials from a Bernoulli process until a predetermined number r of successes has been obtained.

The number of trials needed is a random variable $\tilde{n}$ with a Pascal or Negative binomial distribution

$$P\left(\tilde{n}=n \mid r, p\right)=\binom{n-1}{r-1} p^{r}\left(1-p\right)^{n-r} \quad , n=r, r+1, \ldots$$

Special case, when r = 1: *First success (Fs) distribution*

$$P\left(\tilde{n}=n \mid p\right)=p\left(1-p\right)^{n-1} \quad , n=1,2, \ldots$$

Related to the *Geometric distribution*

$$P\left(\tilde{x}=x \mid p\right)=p\left(1-p\right)^{x} \quad , x=0,1, \ldots$$

- The Poisson process

A counting process with so-called *independent increments*

The events to be counted appears with an intensity $\lambda(t)$

The number of events appearing in the time interval (t_1, t_2) is Poisson distributed with mean

$$\mu = \int_{t_1}^{t_2} \lambda(t) dt$$

i.e

$$P\left(\tilde{r} = r \middle| \lambda(t), t_1, t_2\right) = \frac{\left(\int_{t_1}^{t_2} \lambda(t) dt\right)^r e^{-\int_{t_1}^{t_2} \lambda(t) dt}}{r!} \quad , r = 0,1,\ldots$$

Most common case: $\lambda(t) \equiv \lambda$ (constant) and $t_1 = 0$. $t_2 = t$ (homogeneous process):

$$P\left(\tilde{r} = r \middle| \lambda, t\right) = \frac{(\lambda t)^r e^{-\lambda t}}{r!} \quad , r = 0,1,\ldots$$

Bayes' theorem applied to discrete probability distributions

What is observed is what (normally) has a discrete probability distribution.

- In Binomial sampling we observe the number of successes
- In Hypergeometric sampling we observe the number of success items in the sample
- In Pascal sampling we observe how many items need to be sampled
- In a counting experiment we count the number of events in a specified time interval

Hence, the discrete probability distribution applicable rules the *likelihood*.

Bayes' theorem on a very generic form:

$$P(\theta \mid \text{Data}) \propto L(\theta; \text{Data}) \cdot P(\theta)$$

where P is the probability measure applicable to the parameter θ and $L(\theta; \text{Data})$ is the likelihood of θ in light of the observed Data.

Proportionality constant: $\int_\theta L(\theta; \text{Data})\, dP(\theta) = \langle \text{often} \rangle = \int_\theta L(\theta; \text{Data}) \cdot P(\theta)\, d\theta$

Hence,

for binomial sampling $$P(p \mid n, r) \propto \binom{n}{r} p^r (1-p)^{n-r} \cdot P(p)$$

for hypergeometric sampling $$P(p \mid N, n, r) \propto \left[\binom{Np}{r} \binom{N(1-p)}{n-r} \Big/ \binom{N}{n} \right] \cdot P(p)$$

for Pascal sampling $$P(p \mid n, r) \propto \binom{n-1}{r-1} p^r (1-p)^{n-r} \cdot P(p)$$

for counting in a homogeneous Poisson process $$P(\lambda \mid r, t) \propto \frac{(\lambda t)^r e^{-\lambda t}}{r!} \cdot P(\lambda)$$

Extending with *hyper parameters* ($\boldsymbol{\psi}$)

$$P\left(\theta \mid \text{Data}, \boldsymbol{\psi}\right) \propto L\left(\theta; \text{Data}\right) \cdot P\left(\theta \mid \boldsymbol{\psi}\right)$$

When θ is continuous and the probability measure is Riemann-Stieltjes integrable (there is a cumulative distribution function)

$$f\left(\theta \mid \text{Data}, \boldsymbol{\psi}\right) \propto L\left(\theta; \text{Data}\right) \cdot f\left(\theta \mid \boldsymbol{\psi}\right)$$

where *f* stands for a *probability density function* (its form may very well depend on the conditions ($\boldsymbol{\psi}$ and ($\boldsymbol{\psi}$, Data) respectively)

Exercise 3.24

You feel that $\tilde{p}$, the probability of heads on a toss of a particular coin is either 0.4, 0.5 or 0.6. Your prior probabilities are $P(0.4) = 0.1$, $P(0.5) = 0.7$ and $P(0.6) = 0.2$. You toss the coin three times and obtain heads once and tails twice. What are the posterior probabilities? If you then toss the coin three *more* times and once again obtain heads once and tails twice, what are the posterior probabilities? Also, compute the posterior probabilities by pooling the two samples and revising the original probabilities just once; compare with your previous answers.

Likelihoods from first sample:

$$L(p = 0.4; \text{First}) = \binom{3}{1} \cdot 0.4^1 \cdot 0.6^2 = 0.432$$

$$L(p = 0.5; \text{First}) = \binom{3}{1} \cdot 0.5^1 \cdot 0.5^2 = 0.375$$

$$L(p = 0.6; \text{First}) = \binom{3}{1} \cdot 0.6^1 \cdot 0.4^2 = 0.288$$

Posterior probabilities:

$$P(p \mid \text{First}) = \frac{L(p;\text{First}) \cdot P(p)}{L(0.4;\text{First}) \cdot P(0.4) + L(0.5;\text{First}) \cdot P(0.5) + L(0.6;\text{First}) \cdot P(0.6)}$$

$$\Rightarrow$$

$$P(0.4 \mid \text{First}) = \frac{L(0.4;\text{First}) \cdot P(0.4)}{L(0.4;\text{First}) \cdot P(0.4) + L(0.5;\text{First}) \cdot P(0.5) + L(0.6;\text{First}) \cdot P(0.6)}$$

$$= \frac{0.432 \cdot 0.1}{0.432 \cdot 0.1 + 0.375 \cdot 0.7 + 0.288 \cdot 0.2} = 0.11891$$

$$P(0.5 \mid \text{First}) = \frac{0.375 \cdot 0.7}{0.432 \cdot 0.1 + 0.375 \cdot 0.7 + 0.288 \cdot 0.2} = 0.7225434$$

$$P(0.6 \mid \text{First}) = \frac{0.288 \cdot 0.2}{0.432 \cdot 0.1 + 0.375 \cdot 0.7 + 0.288 \cdot 0.2} = 0.1585467$$

Likelihoods from second sample:

$$L(p=0.4;\text{Second}) = \binom{3}{1} \cdot 0.4^1 \cdot 0.6^2 = 0.432$$

$$L(p=0.5;\text{Second}) = \binom{3}{1} \cdot 0.5^1 \cdot 0.5^2 = 0.375$$

$$L(p=0.6;\text{Second}) = \binom{3}{1} \cdot 0.6^1 \cdot 0.4^2 = 0.288$$

These are the same values as with the first sample since the sample outcomes are identical.

Posterior probabilities after second sample;

$$P\left(p \middle| \text{Second}, (\text{First})\right)$$

$$= \frac{L\left(p;\text{Second}\right)\cdot P\left(p \middle| \text{First}\right)}{L\left(0.4;\text{Second}\right)\cdot P\left(0.4 \middle| \text{First}\right) + L\left(0.5;\text{Second}\right)\cdot P\left(0.5 \middle| \text{First}\right) + L\left(0.6;\text{Second}\right)\cdot P\left(0.6 \middle| \text{First}\right)}$$

$$P\left(0.4 \middle| \text{Second}, (\text{First})\right)$$

$$= \frac{L\left(0.4;\text{Second}\right)\cdot P\left(0.4 \middle| \text{First}\right)}{L\left(0.4;\text{Second}\right)\cdot P\left(0.4 \middle| \text{First}\right) + L\left(0.5;\text{Second}\right)\cdot P\left(0.5 \middle| \text{First}\right) + L\left(0.6;\text{Second}\right)\cdot P\left(0.6 \middle| \text{First}\right)}$$

$$= \frac{0.432\cdot 0.11891}{0.432\cdot 0.11891 + 0.375\cdot 0.7225434 + 0.288\cdot 0.1585467} = 0.1395959$$

$$P\left(0.5 \middle| \text{Second}, (\text{First})\right)$$

$$= \frac{0.375\cdot 0.7225435}{0.432\cdot 0.11891 + 0.375\cdot 0.7225434 + 0.288\cdot 0.1585467} = 0.7363188$$

$$P\left(0.6 \middle| \text{Second}, (\text{First})\right)$$

$$= \frac{0.288\cdot 0.1585467}{0.432\cdot 0.11891 + 0.375\cdot 0.7225434 + 0.288\cdot 0.1585467} = 0.1240853$$

Likelihoods from pooled samples:

$$L\left(p=0.4;\text{Pooled}\right)=\binom{6}{2}\cdot 0.4^2\cdot 0.6^4=0.31104$$

$$L\left(p=0.5;\text{Pooled}\right)=\binom{6}{2}\cdot 0.5^2\cdot 0.5^4=0.234375$$

$$L\left(p=0.6;\text{Pooled}\right)=\binom{6}{2}\cdot 0.6^2\cdot 0.4^4=0.13824$$

Posterior probabilities:

$$P\left(p \middle| \text{Pooled}\right) = \frac{L\left(p;\text{Pooled}\right)\cdot P\left(p\right)}{L\left(0.4;\text{Pooled}\right)\cdot P\left(0.4\right) + L\left(0.5;\text{Pooled}\right)\cdot P\left(0.5\right) + L\left(0.6;\text{Pooled}\right)\cdot P\left(0.6\right)}$$

$$\Rightarrow$$

$$P\left(0.4 \middle| \text{Pooled}\right) = \frac{L\left(0.4;\text{Pooled}\right)\cdot P\left(0.4\right)}{L\left(0.4;\text{Pooled}\right)\cdot P\left(0.4\right) + L\left(0.5;\text{Pooled}\right)\cdot P\left(0.5\right) + L\left(0.6;\text{Pooled}\right)\cdot P\left(0.6\right)}$$

$$= \frac{0.31104\cdot 0.1}{0.31104\cdot 0.1 + 0.234375\cdot 0.7 + 0.13824\cdot 0.2} = 0.1395959$$

$$P\left(0.5 \middle| \text{Pooled}\right)$$

$$= \frac{0.234375\cdot 0.7}{0.31104\cdot 0.1 + 0.234375\cdot 0.7 + 0.13824\cdot 0.2} = 0.7363188$$

$$P\left(0.6 \middle| \text{Pooled}\right)$$

$$= \frac{0.13824\cdot 0.2}{0.31104\cdot 0.1 + 0.234375\cdot 0.7 + 0.13824\cdot 0.2} = 0.1240853$$

Comparison:

$$P(0.4 \mid \text{Second}, (\text{First})) = 0.1395959 \qquad P(0.4 \mid \text{Pooled}) = 0.1395959$$

$$P(0.5 \mid \text{Second}, (\text{First})) = 0.7363188 \qquad P(0.5 \mid \text{Pooled}) = 0.7363188$$

$$P(0.6 \mid \text{Second}, (\text{First})) = 0.1240853 \qquad P(0.6 \mid \text{Pooled}) = 0.1240853$$

Identical results! Expected?

$$P(p \mid \text{Second}, (\text{First}))$$

$$= \frac{L(p;\text{Second}) \cdot P(p \mid \text{First})}{L(0.4;\text{Second}) \cdot P(0.4 \mid \text{First}) + L(0.5;\text{Second}) \cdot P(0.5 \mid \text{First}) + L(0.6;\text{Second}) \cdot P(0.6 \mid \text{First})}$$

$$L(p;\text{Second}) \cdot P(p \mid \text{First})$$

$$= L(p;\text{Second}) \cdot \frac{L(p;\text{First}) \cdot P(p)}{L(0.4;\text{First}) \cdot P(0.4) + L(0.5;\text{First}) \cdot P(0.5) + L(0.6;\text{First}) \cdot P(0.6)}$$

$$\propto L(p;\text{Second}) \cdot L(p;\text{First}) \cdot P(p) = \binom{3}{1} p^1 (1-p)^2 \binom{3}{1} p^1 (1-p)^2 \cdot P(p)$$

$$\propto p^2 (1-p)^4 \cdot P(p)$$

$$P\left(p \mid \text{Pooled}\right) = \frac{L\left(p;\text{Pooled}\right)\cdot P\left(p\right)}{L\left(0.4;\text{Pooled}\right)\cdot P\left(0.4\right) + L\left(0.5;\text{Pooled}\right)\cdot P\left(0.5\right) + L\left(0.6;\text{Pooled}\right)\cdot P\left(0.6\right)}$$

$$\propto \binom{6}{2} p^2\left(1-p\right)^4 \cdot P\left(p\right) \propto p^2\left(1-p\right)^4 \cdot P\left(p\right)$$

Hence the two ways of computing the posterior probabilities using both samples are always identical

Exercise 3.33

Suppose that you feel that accidents along a particular stretch of highway occur roughly according to a Poisson process and that the intensity of the process is either 2, 3 or 4 accidents per week. Your prior probabilities for these three possible intensities are 0.25, 0.45 and 0.30, respectively. If you observe the highway for a period of three weeks and 10 accidents occur, what are your posterior probabilities?

Likelihoods:

$$L(\lambda = 2; r = 10, t = 3) = \frac{(\lambda \cdot t)^r e^{-\lambda \cdot t}}{r!} = \frac{(2 \cdot 3)^{10} e^{-2 \cdot 3}}{10!} = 0.04130309$$

$$L(\lambda = 3; r = 10, t = 3) = \frac{(3 \cdot 3)^{10} e^{-3 \cdot 3}}{10!} = 0.1185801$$

$$L(\lambda = 4; r = 10, t = 3) = \frac{(4 \cdot 3)^{10} e^{-4 \cdot 3}}{10!} = 0.1048373$$

Posterior probabilities:

$$P\left(\lambda \middle| r=10, t=3\right)=\frac{\frac{(3\lambda)^{10} e^{-3\lambda}}{10!}\cdot P(\lambda)}{\frac{(3\cdot 2)^{10} e^{-3\cdot 2}}{10!}\cdot P(2)+\frac{(3\cdot 3)^{10} e^{-3\cdot 3}}{10!}\cdot P(3)+\frac{(3\cdot 4)^{10} e^{-3\cdot 4}}{10!}\cdot P(4)}$$

$$\left\langle \begin{array}{l}\text{Faculties}\\ \text{and } 3^{10} \text{ terms}\\ \text{cancel out}\end{array}\right\rangle = \frac{\lambda^{10} e^{-3\lambda}\cdot P(\lambda)}{2^{10} e^{-3\cdot 2}\cdot P(2)+3^{10} e^{-3\cdot 3}\cdot P(3)+4^{10} e^{-3\cdot 4}\cdot P(4)}$$

$$P\left(\lambda=2 \middle| r=10, t=3\right)=\frac{2^{10} e^{-3\cdot 2}\cdot 0.25}{2^{10} e^{-3\cdot 2}\cdot 0.25+3^{10} e^{-3\cdot 3}\cdot 0.45+4^{10} e^{-3\cdot 4}\cdot 0.30}=0.1085347$$

$$P\left(\lambda=3 \middle| r=10, t=3\right)=\frac{3^{10} e^{-3\cdot 3}\cdot 0.45}{3^{10} e^{-3\cdot 2}\cdot 0.25+3^{10} e^{-3\cdot 3}\cdot 0.45+4^{10} e^{-3\cdot 4}\cdot 0.30}=0.5608804$$

$$P\left(\lambda=4 \middle| r=10, t=3\right)=\frac{4^{10} e^{-3\cdot 4}\cdot 0.30}{2^{10} e^{-3\cdot 2}\cdot 0.25+3^{10} e^{-3\cdot 3}\cdot 0.45+4^{10} e^{-3\cdot 4}\cdot 0.30}=0.3305849$$

Predictive distributions

For an unknown parameter of interest, θ, we would – according to the subjective interpretation of probability

- assign a prior distribution
- upon obtaining data related to θ, compute a posterior distribution

The prior and posterior distributions are used to *make inference* about the unknown θ – *explanatory inference*

We may also be interested in *predictive inference* , i.e. predict data related to θ not yet obtained

For cross-sectional data the term prediction is mostly used, while for time series data we rather use the term *forecasting*.

Let $y_1, \ldots, y_M, \ldots$ be the set (finite or infinite) of observed values we may obtain under conditions ruled by the unknown θ.

The uncertainty associated with each observation – i.e. that its value/state cannot be known in advance – is modelled by letting the observed value be the realisation of a random variable $\tilde{y}$ with a probability distribution depending on θ:

$$P(\tilde{y} = y_k | \theta) = f(y_k | \theta)$$

Prior-predictive distributions

The prior-predictive distribution of $\tilde{y}$ is the set of marginal probabilities obtained when the dependency on θ is integrated/summed out by weighting the probability mass function $f(y|\theta)$ with the prior distribution of θ:

$$P(\tilde{y} = y_k) = \begin{cases} \sum_\theta f(y_k|\theta) \cdot P(\tilde{\theta} = \theta) & \text{if } \theta \text{ assumes a enumerable set of values} \\ \int_\theta f(y_k|\theta) \cdot p(\theta) d\theta & \text{if } \theta \text{ assumes values on a continuous scale} \end{cases}$$

Posterior-predictive distributions

The posterior-predictive distribution of $\tilde{y}$ is the set of marginal probabilities obtained when the dependency on θ is integrated/summed out by weighting the probability mass function $f(y|\theta)$ with the posterior distribution of θ given an already obtained set of observations (Data):

$$P\left(\tilde{y} = y_k \middle| \text{Data}\right)$$

$$= \begin{cases} \sum_\theta f\left(y_k \middle| \theta\right) \cdot P\left(\tilde{\theta} = \theta \middle| \text{Data}\right) & \text{if } \theta \text{ assumes a enumerable set of values} \\ \int_\theta f\left(y_k \middle| \theta\right) \cdot p\left(\theta \middle| \text{Data}\right) d\theta & \text{if } \theta \text{ assumes values on a continuous scale} \end{cases}$$