Bayesian Modelling

Philosophical view

1. Bayesian modelling is different from traditional parametric in a key way, we are trying to model an event, and the parameter used in modelling the event. To do this, we form a prior distribution of our parameter that we believe will be accurate in modelling the underlying event. We observe the event and then our goal is to find the posterior.

2. the posterior is defined as: $f(\theta \,|\, \text{event} = \text{event})$ expanding this, we get

$$f(\theta|event = event) = \frac{f(\theta) \times p(Event|\theta)}{P(Event)} \tag{1}$$

3. The β distribution with parameters a and b is a distribution often used in modelling priors, the pdf takes the form:

$$f_{a,b} = \frac{t^{a-1}(1-t)^{b-1}}{\beta(a,b)}$$
$$\text{where } \beta = \int_u^\infty u^{a-1}(1-u)^{b-1}du \tag{2}$$

4. Lastly, the β distribution is a conjugate prior, meaning if we use a β distribution to model the prior of our parameter, the posterior will also be a β distribution with a different parameter which is mathematically convenient.

Expectation and Variance of Random Variables

Expectation: The expectation or mean of a random variable $\tilde{a}$ is equal to:

$$E(\tilde{a}) = \frac{1}{n} \times \sum_{i=1}^n \tilde{a}_i p_{\tilde{a}}(\tilde{a} = \tilde{a}_i)$$

Linearity of Expectation: The expectation or mean of the sum of two random variable $\tilde{a}$ and $\tilde{b}$ is equal to:

$$E(\tilde{a} + \tilde{b}) = E(\tilde{a}) + E(\tilde{b})$$

Mean of scaled Random Variables: The expectation or mean of a scaled random variable $\alpha\tilde{a}$ is equal to:

$$E(\alpha\tilde{a}) = \alpha E(\tilde{a})$$

Multiplication of Random Variables:
If two random variables, $\tilde{a}$ and $\tilde{b}$ come from the same distribution and are independent then the following equality holds:

$$E(\tilde{a}\tilde{b}) = E^2(\tilde{a}) = E^2(\tilde{b}) = E(\tilde{a})E(\tilde{b})$$

Variance: The variance of a random variable $\tilde{a}$ is equal to:

$$Var(\tilde{a}) = E((\tilde{a} - \mu_{\tilde{a}})^2) = \frac{1}{n} \times \sum_{i=1}^{n} (\tilde{a}_i - \mu_{\tilde{a}})^2$$

We can use a shortcut to express the variance in a simplified form:

$$\begin{aligned}
Var(\tilde{a}) &= E((\tilde{a} - \mu_{\tilde{a}})^2) \\
&= E(\tilde{a}^2 - 2\mu_{\tilde{a}}\tilde{a} + \mu_{\tilde{a}}^2) \\
&= E(\tilde{a}^2) - \mu_{\tilde{a}}E(2\tilde{a} + \mu_{\tilde{a}}) \\
&= E(\tilde{a}^2) - 2\mu_{\tilde{a}}^2 + \mu_{\tilde{a}}^2 \\
&= E(\tilde{a}^2) - E^2(\tilde{a})
\end{aligned} \tag{3}$$

Variance of the sum of two random variables:

$$\begin{aligned}
Var(\tilde{a} + \tilde{b}) &= E((\tilde{a} + \tilde{b} - E(\tilde{a} + \tilde{b}))^2) \\
&= E((\tilde{a} + \tilde{b} - \mu\tilde{a} - \mu_{\tilde{b}}))^2) \\
&= E((\tilde{a} - \mu_{\tilde{a}})^2) + E((\tilde{b} - \mu_{\tilde{b}})^2) + 2E((\tilde{a} - \mu_{\tilde{a}})(\tilde{b} - \mu\tilde{b})) \\
&= Var(\tilde{a}) + Var(\tilde{b}) + 2Cov(\tilde{a}, \tilde{b})
\end{aligned} \tag{4}$$

If the two random variables are uncorrelated then:

$$Var(\tilde{a} + \tilde{b}) = Var(\tilde{a}) + Var(\tilde{b})$$

PDF, PMF Mean and Variance of Popular Distributions:

Binomial:

$$Mean = n\theta \qquad Variance = n\theta(1 - \theta) \qquad PMF = \binom{n}{k} \theta^k (1 - \theta)^{n-k}$$

Bernoulli:

$$Mean = \theta \qquad Variance = \theta(1 - \theta) \qquad PMF = \begin{cases} \theta & if\, k = 1 \\ 1 - \theta & if\ k = 0 \end{cases}$$

Geometric:

$$Mean = \frac{1}{\theta} \qquad Mean\,Squared = E(\tilde{x}^2) = \frac{2 - \theta}{\theta^2} \qquad Variance = \frac{1 - \theta}{\theta^2} \qquad PMF = (1-\theta)^{k-1}\theta$$

Exponential:

$$Mean = E(\tilde{x}) = \frac{1}{\lambda} \qquad Mean\,Squared = E(\tilde{x}^2) = \frac{2}{\lambda^2} \qquad Variance = \frac{1}{\lambda^2} \qquad PMF = \lambda e^{-\lambda t}$$

Poisson:

$$Mean = E(\tilde{x}) = \lambda \qquad Mean\,Squared = E(x^2) = \lambda + \lambda^2 \qquad Variance = \lambda \qquad PMF = (1-\theta)^{k-1}\theta$$

Gaussian:

$$Mean = \mu \qquad Variance = \sigma^2$$

Uniform:

$$Mean = \frac{a+b}{2} \qquad Variance = \frac{(b-a)^2}{12} \qquad PDF = \frac{1}{b-a}$$

Distribution	Parameter	Maximum-likelihood estimator	Mean
Bernoulli	θ	$\frac{1}{n}\sum_{i=1}^{n} x_i$	θ
Geometric	θ	$\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^{-1}$	θ^{-1}
Poisson	λ	$\frac{1}{n}\sum_{i=1}^{n} x_i$	λ
Exponential	λ	$\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^{-1}$	λ^{-1}
Gaussian	μ	$\frac{1}{n}\sum_{i=1}^{n} x_i$	μ

Joint, Marginal, and Conditional Distributions

Joint

  Marginal If you want the marginal of x, you sum over y **Conditional**

$$Conditional = \frac{joint}{marginal}$$

Standardized Random Variables and Linear Estimation

Standardizing a Random Variable: How to standardize a random variable, $\tilde{a}$, such that its mean centered and unit variance:

$$s(a) = \frac{\tilde{a} - \mu_{\tilde{a}}}{\sigma_{\tilde{a}}}$$

Linear Estimates: Best linear estimate of b given a is correlation coefficient times a equals b estimate (plus error):

$$\hat{b} = \mu_{\tilde{b}} + \sigma_{\tilde{b}} \times \rho_{\tilde{a},\tilde{b}} \times s(\tilde{a})$$

Remember, if $\tilde{a}$ isn't standardized, make sure to standardize it. For hw/test problems, start by defining what you're predicting, and what the predictor is, then calculate the expectation, variance, and std. dev of your predictor, the cov and corr of the predictor/predicion, then fill in formula.

If our random variables $\tilde{a}$ and $\tilde{b}$ are standardized, the linear estimate is expressed in the following way:

$$\hat{b} = \rho_{\tilde{a},\tilde{b}} \times s(\tilde{a})$$

Which makes sense as

$$E(\tilde{b}) = 0 \text{ and } Var(\tilde{b}) = 1$$

Co-Variance and Correlation

Covariance: We define the co-variance of $\tilde{a}$ and $\tilde{b}$:

$$Cov(\tilde{a}, \tilde{b}) = E((\tilde{a} - \mu_{\tilde{a}})(\tilde{b} - \mu\tilde{b}))$$

We can use a shortcut to express the co-variance in a simplified form:

$$\begin{aligned}
Cov(\tilde{a}, \tilde{b}) &= E((\tilde{a} - \mu_{\tilde{a}})(\tilde{b} - \mu\tilde{b})) \\
&= E(\tilde{a}\tilde{b}) - \mu_{\tilde{b}}E(\tilde{a}) - \mu_{\tilde{a}}E(\tilde{b}) + \mu_{\tilde{a}}\mu_{\tilde{b}} \\
&= E(\tilde{a}\tilde{b}) - \mu_{\tilde{a}}\mu_{\tilde{b}} - \mu_{\tilde{a}}\mu_{\tilde{b}} + \mu_{\tilde{a}}\mu_{\tilde{b}} \\
&= E(\tilde{a}\tilde{b}) - \mu_{\tilde{a}}\mu_{\tilde{b}}
\end{aligned} \tag{5}$$

If our random variables $\tilde{a}$ and $\tilde{b}$ are standardized, correlation is expressed the following way:

$$Cov(\tilde{a}, \tilde{b}) = E(\tilde{a}\tilde{b})$$

Which makes sense as:

$$E(\tilde{a}) = E(\tilde{b}) = 0$$

Correlation: The correlation between two random variables $\tilde{a}$ and $\tilde{b}$ is defined as:

$$Corr(\tilde{a}, \tilde{b}) = \rho_{\tilde{a},\tilde{b}} = \frac{Cov(\tilde{a}, \tilde{b})}{\sigma_{\tilde{a}} \times \sigma_{\tilde{b}}}$$

If the random variables are mean centered and standardized then the correlation is the same as co-variance:

$$Corr(\tilde{a}, \tilde{b}) = \rho_{\tilde{a}, \tilde{b}} = Cov(\tilde{a}, \tilde{b})$$

Which makes sense as

$$\sigma_{\tilde{a}}^2 = 0 \text{ and } \sigma_{\tilde{b}}^2 = 0$$

Mean Squared Error of MMSE Estimate: The linear MMSE estimate of $\tilde{b}$ given $\tilde{a}$ is:

$$E((\alpha \times \tilde{a} + \beta - \tilde{b})^2) = (1 - \rho_{\tilde{a}, \tilde{b}}^2) \times \sigma_{\tilde{b}}^2$$

Consequences of the MMSE Estimate:

1. correlation coefficient is bounded between -1 and 1.

2. When it equals -1 or 1, it implies a linear relationship between the random variables $\tilde{a}$ and $\tilde{b}$

3. The square of the correlation coefficient is the fraction of variance in $\tilde{b}$ that can be explained linearly in terms of $\tilde{a}$

Main Takeaways

Facts about correlation, expectation, co-variance, and standardized random variables:

1. If your random variables are mean centered and standardized, then the mean is 0 and the variance is 1. The variance will also equal the expectation of your random variable squared.

   Expectation $\rightarrow E(\tilde{a}) = 0$ Variance $\rightarrow E(\tilde{a}^2) = 1$

2. If you have two (or more) random variables that are mean-centered and standardized, the correlation equals to covariance

   $Cov(\tilde{a}, \tilde{b}) = Corr(\tilde{a}, \tilde{b})$

3. Linear Estimate residuals have an expectation of 0

   $E(\tilde{b} - (\mu_{\tilde{b}} + \sigma_{\tilde{b}} \times s(\tilde{a}))) = 0$

4. Linear Estimates and residuals have 0 co-variance, and are uncorrelated

5. Two random variables that are independent implies that they are also uncorrelated. Surprisingly, the opposite is not true. Two random variables being uncorrelated does not mean they are independent.

Problem Solving Strategies

Inverse Sampling

1. Step 1: Determine whether the problem is asking you to find the X that would result in a uniform distribution value, or find the x that would give you the supplied uniform distribution.

2. Case 1: In the case you're given a uniform distribution value, and want to find the corresponding x.

   calculate the CDF of the function

   Swap the x and y in the CDF and solve for y again, obtaining the inverse cdf

   If your CDF consists of multiple functions depending on the value of x, determine which inverse CDF you should use by looking at the given value of the uniform distribution sample.

   plug in the uniform value as x into the proper inverse cdf, solving for y

3. Case 2: If you are given an x value and trying to get the corresponding uniform value.

   plug x into the appropriate CDF and you will get a value in the 0-1 domain

Markov's Inequality

Let $\tilde{a}$ be a non-negative random variable. For any positive constant $c > 0$

$$P(\tilde{a} \geq c) \leq \frac{E(\tilde{a})}{c} \tag{6}$$

Chebyshev's Inequality

For any positive constant $c > 0$ and any random variable $\tilde{a}$ with bounded variance:

$$P((\tilde{a} - E(\tilde{a}) \geq c)) \leq \frac{Var(\tilde{a})}{c} \tag{7}$$

Chebyshev Continued

Probability that the sample mean deviates from the population mean by more than c (Where n is sample size).

$$P(\bar{x} - \mu > c) = \frac{\sigma^2}{nc^2} \tag{8}$$

Step 1: If you don't know the variance, make a reasonable estimate of the upper bound.
Step 2: If you're given a bound with respect to distance for your confidence interval,(i.e. we want to have a 95% confidence interval that does not exceed 5cm), you'll have to divide the

distance (width) by 2, and put that value into Chevysevs.
Step 3: Do the following computation:

$$P(\bar{x} - \mu > c) = 1 - \alpha = \frac{\sigma^2}{n(\frac{c}{2})^2}$$

Step 4: The answer from the computation gives us the minimum sample size we need to guarantee a confidence interval of a certain width.

Confidence Interval

Take the mean, and get your bounds by adding (for top bound) and subtracting (for lower bound) the following expression:

$$Upper\ Bound = \mu + \frac{\sigma}{\sqrt{n}} \times \text{z-score} \qquad Lower\ Bound = \mu - \frac{\sigma}{\sqrt{n}} \times \text{z-score}$$

Popular Z Scores:

$$\begin{pmatrix} ConfidenceLevel & Area\ between\ 0\ and\ z-score & z-score \\ 90\% & 0.4500 & 1.645 \\ 95\% & 0.4750 & 1.960 \\ 98\% & 0.4900 & 2.326 \\ 99\% & 0.4950 & 2.576 \end{pmatrix}$$

Miscellaneous

Gaussian vectors

1. A Gaussian vectors x in R^d has co-variance matrix $x^T x$.

2. The probability density of Gaussian vector x is defined as $x^T \Sigma^{-1} x$ where Σ is the inverse of the co-variance matrix.

3. Any linear transformation of a Gaussian vector results in a Gaussian vector, so the equation:

    As a result the following is a Gaussian vector where A is an $R^m x d$ matrix

    $$y = Ax + b \qquad (9)$$

4. Additionally the marginal and conditional distributions of Gaussian vectors are Gaussian vectors

    the density of the marginal distribution is

    $$f(a) = (1/\sqrt{2\pi}) \times \exp \frac{-a^2}{2} \qquad (10)$$

The conditional density is given by:

$$f(b|a) = (1/\sqrt{2\pi \times (1-\rho^2)}) \times \exp \frac{(b - \rho \times a)^2}{2 \times (1 - \rho^2)} \tag{11}$$

Lastly the mean and variance of a gaussian vector b given a is:

$$\mu_{b|a} = \mu_b \frac{\rho \sigma_b (a - \mu_a)}{\sigma_b} \tag{12}$$

$$\sigma_{b|a}^2 = (1 - \rho^2)\sigma_b^2 \tag{13}$$

Proof: Final Exam Study Question 5

$$Var(n\tilde{x} + \sum_{i=1}^{n} \tilde{z}_i)$$

$$= Var(n\tilde{x} + n \cdot \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i)$$

$$= Var(n(\tilde{x} + \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i))$$

$$= n^2 Var(\tilde{x} + \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i)$$

$$= n^2 (Var(\tilde{x}) + Var(\frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i))$$

given x and z are independent

$$= n^2 (1 + \frac{\sigma^2}{n})$$

$$= n^2 + n\sigma^2$$

$$Var(\frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i)$$

$$= \frac{1}{n^2} Var(\sum_{i=1}^{n} \tilde{z}_i)$$

$$= \frac{1}{n^2} n\sigma^2$$

$$= \frac{\sigma^2}{n}$$