Parametric Categorization

One popular application of Bayesian inference is to make inference based on previously observed items. In this tutorial, we'll implement how to infer parameters of a category distribution based on previous observations of exemplars.

Suppose people process different types of words at different speeds. In an experiment, you presented to participants some normal monosyllabic English words (e.g., cake, jar, shoe), one at a time. Participants pressed a key as soon as they knew what word had been presented.

You obtained the following response time (RTs) measurements on four normal words (in milliseconds):

[594 483 672 491]

So, what is the overall average (or, more technically, the population mean μ) RT for this "normal" category? With all the observations, your best guess should be the sample mean:

$$\bar{x} = \frac{594 + 483 + 672 + 491}{4} = 560$$

By making this guess, you're claiming that "μ = 560". However, how certain are you in making this claim? The number is just an estimate. It doesn't tell you how certain (or uncertain) we are about this estimation. Knowing the uncertainty is helpful, especially in the case of a categorization problem.

If we formulate the problem as a parametric categorization problem, we can assume a functional form for the RTs of this "normal monosyllabic word" category. Let's assume it to be a normal distribution:

$$P(x|\ \mu, \sigma^2) = N(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-(x-\mu)^2/2\sigma^2\right)}$$

, where x is the RT of a normal monosyllabic word, and μ and σ^2 are the two parameters for this distribution, which are the population mean RT and variance, respectively. Let's assume we know that $\sigma^2 = 1600$ from previous literature.

Since our goal is to infer μ based on previous observation, we can formulate this as a Bayesian inference problem: We aim to obtain the **posterior distribution on $\boldsymbol{\mu}$**, given all the observations and the variance, i.e., $P(\mu \mid x_1, x_2 \ldots, x_N, \sigma^2)$, where $x_1, x_2, \ldots x_N$ refer to the observed response times of the N tested words.

We assume that the prior distribution of μ is uniform within the range of 1 to 1500.

We will use both discretize-space method and sampling method to implement this categorization model. We will define the space of μ with the range of 1 to 1500 ms with the stepsize of 1.

in-class Exercise (submit category1.m at CCLE)

Part 1: use discretize-space method to implement the steps 1-3

1. **Compute prior.**
2. **Compute and plot likelihood functions** for each of the observations, $P(x_1|\ \mu, \sigma^2)$, $P(x_2|\ \mu, \sigma^2)$, $P(x_3|\ \mu, \sigma^2)$, $P(x_4|\ \mu, \sigma^2)$.
3. **Compute and plot the posterior distribution** of , $P(\mu|x_1, x_2, x_3, x_4, \sigma^2)$ in the same figure window as the likelihood functions.

4. Use computed posterior distribution of population mean RT μ to **calculate the expected value $E(\mu)$ and variance $Var(\mu)$ of μ** .

Recall the formula for computing the expected value (i.e., mean) from a continuous distribution:

$$E(X) = \int_{-\infty}^{+\infty} x\, P(X = x)dx \qquad \Rightarrow \qquad E(\mu) = \int_{-\infty}^{+\infty} t\, P(\mu = t)dt$$

As we're approximating the continuous distribution with a discrete array, we'll do the Riemann sum:

$$E(\mu) \approx \sum_{i=1}^{N} \mu_i P(\mu = \mu_i)\, \Delta\mu \tag{1}$$

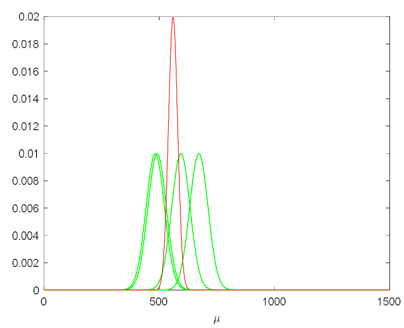
, where $\Delta\mu$ is the step size in the μ vector, and N, in our case, is 1500.

The variance can also be approximated similarly, after computing the expected value.

The formula for variance of a continuous random variable X is:

$$Var(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2\, P(X = x)dx \tag{2}$$

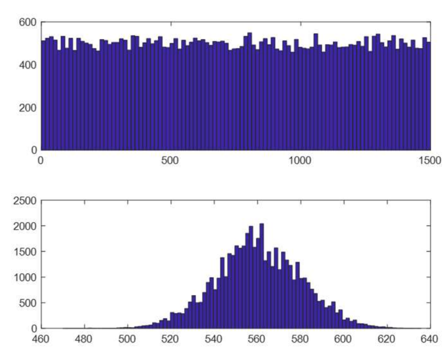
Based on the equation 2, use Riemann sum approximation to compute an approximated variance $Var(\mu)$ for the posterior distribution of μ.




```
PostMean =

        560


PostVar =

        400
```

Part 2: use sampling method to implement the steps 5-6

5. Implement the Bayesian model **using sampling method**, and use "hist" function to **plot the histograms from prior samples and posterior samples.**
6. **Compute the mean and variance from posterior samples**. I used sample size 50000 for the simulation. If your computer is slow with simulation, you can reduce the sample size to 5000 or 10000.




```
PostMeanSample =

     560.31


PostVarSample =

     396.62
```