IIT Jodhpur

Biological Vision and Applications
Module 03-05: Parameter Estimation

Hiranmay Ghosh

How to estimate a parameter ?

Maximum Likelihood estimation

- Bayesian framework of reasoning assumes some conditional probabilities (priors)
  - e.g., $P(\text{Red} \mid \text{Banana}) = 0.1$
- Where do you get the number from?
  - Domain theory
  - Past observations
- From your past observations
  - You observed 20 bananas; 2 of them are red
  - $P(\text{red} \mid \text{banana}) = \frac{2}{20} = 0.1$

- **Maximum Likelihood Estimation (MLE)**
  - Let X be a stochastic variable with n possible values: $x_1, \ldots, x_n$
  - We make N experiments
    - Observe r_i occurences for $X = x_i$ $\quad (\sum_{i=1}^{n} r_i = N)$
  - Estimates of $P(X = x_i) = \frac{r_i}{N}$

Maximum Likelihood estimation

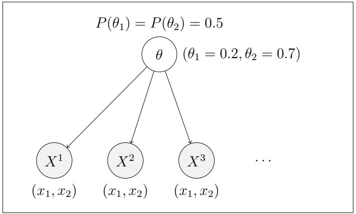
Pitfalls

- Data-driven approach

- Sparsity of data
- Extremely unreliable, if the sample size is small
  - Observe 2 bananas, one of them is red: $P(red \mid banana) = \frac{1}{2}$
  - Intuitively, this is incorrect!

- Does not tell how reliable the estimate is
  - Does not distinguish between (2 out of 20) and (200 out of 2000)
  - Cannot tell $P(X = x_1) = 0.1 \pm 5\%$

Bayesian Estimation



- Assume that X can have two values: (x_1, x_2)
- Hidden parameter $\theta \equiv P(X = x_1)$:
  - Causes outcomes of the experiments
  - $P(X = x_2) = 1 - \theta$
- Given θ, experiments are conditionally independent
- Assume that θ can have two possible values: $(\theta_1 = 0.2, \theta_2 = 0.7)$
- Assume that they are equiprobable to begin with:
  - $P(\theta_1) = P(\theta_2) = 0.5$ $\quad [P(\theta_1) \equiv P(\theta = \theta_1), P(\theta_2) \equiv P(\theta = \theta_2)]$

$$\begin{aligned} P(X^1 = x_1) &= P(\theta_1).P(x_1 \mid \theta_1) + P(\theta_2).P(x_1 \mid \theta_2) \\ &= P(\theta_1).\theta_1 + P(\theta_2).\theta_2 = 0.5 \times 0.7 + 0.5 \times 0.2 = 0.45 \end{aligned}$$

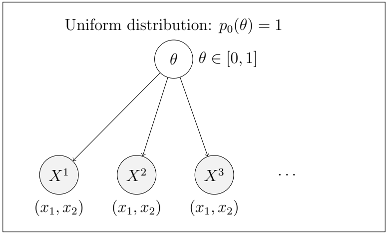
Now, we make the first experiment. Assume $X^1 = x_1$

$$P(\theta_1 \mid X^1 = x_1) = \frac{P(\theta_1).P(X^1 = x_1 \mid \theta_1)}{P(X^1 = x_1)} = \frac{0.5 \times 0.2}{0.45} \approx 0.22$$

$$P(\theta_2 \mid X^1 = x_1) = \frac{P(\theta_2).P(X^1 = x_1 \mid \theta_2)}{P(X^1 = x_1)} = \frac{0.5 \times 0.7}{0.45} \approx 0.78$$

Bayesian Estimation

contd.



Predicted outcome of 1st experiment:

$$\hat{\theta}_0 = P(X^1 = x_1 \mid p_0(\theta)) = \int_0^1 p_0(\theta).\theta.d\theta = \int_0^1 \theta.d\theta = \frac{1}{2}$$

We perform the first experiment. We observe $X_1 = x_1$

$$p_1(\theta) = p(\theta \mid p_0(\theta), X^1 = x_1) = \frac{p_0(\theta).P(X^1=x_1|\theta)}{P(X^1=x_1)} = \frac{1\times\theta}{1/2} = 2\theta$$

Predicted outcome of 2nd experiment:

$$\hat{\theta}_1 = P(X^2 = x_1 \mid p_1(\theta)) = \int_0^1 p_1(\theta).\theta.d\theta = \int_0^1 2.\theta^2.d\theta = \frac{2}{3}$$

Now we make the 2nd experiment. We observe $X^2 = x_2$

$$p_2(\theta) = p(\theta \mid p_1(\theta), X^2 = x_2) = \frac{p_1(\theta).P(X^2=x_2|\theta)}{P(X^2=x_2)} = \frac{2\theta\times(1-\theta)}{1-2/3} = 6.\theta.(1-\theta)$$

Predicted outcome of 3rd experiment:

$$\hat{\theta}_2 = P(X^3 = x_1 \mid p_2(\theta)) = \int_0^1 p_2(\theta).\theta.d\theta = \int_0^1 \frac{2}{3}.\theta^2(1-\theta).d\theta = \frac{1}{2}$$

We could also compute the result in a single step

▷ sequence of observation does not matter

$$p_2(\theta) = p(\theta \mid p_0(\theta), X^1 = x_1, X^2 = x_2) = \frac{p_0(\theta).P(X^1=x_1,X^2=x_2|\theta)}{P(X^1=x_1,X^2=x_2)} = \frac{\theta\times(1-\theta)}{\int_0^1 \theta.(1-\theta).d\theta} = 6.\theta.(1-\theta)$$

P(x) is probability value (discrete variable), *p*(x) is probability density function (continuous variable)

Bayesian Estimation

- Assume that we have made n experiments
  - $D = \langle x_1, x_2, x_1, x_1, \ldots \rangle$
  - We have observed k x_1s and $(n-k)$ x_2s (in whatever sequence)

$$P(D \mid \theta) = \left[{n \atop k}\right].\theta^k.(1-\theta)^{n-k}$$

$$p_n(\theta) = p(\theta \mid p_0(\theta), D) = \frac{p_0(\theta).P(D|\theta)}{P(D)} = \frac{1 \times \left[{n \atop k}\right].\theta^k.(1-\theta)^{n-k}}{\int_0^1 \left[{n \atop k}\right]\theta^k.(1-\theta)^{n-k}.d\theta} = \frac{(n+1)!}{k!.(n-k)!}.\theta^k.(1-\theta)^{n-k}$$

$$\hat{\theta}_n = P(X_{n+1} \mid p_n(\theta)) = \int_0^1 \theta.p_n(\theta).d\theta = \frac{(n+1)!}{k!.(n-k)!}.\int_0^1 \theta^{k+1}.(1-\theta)^{n-k}.d\theta$$

$$= \frac{(n+1)!}{k!.(n-k)!} \times \frac{(k+1)!.(n-k)!}{(n+2)!} = \frac{k+1}{n+2}$$

We have used the result $\int_0^1 x^m.(1-x)^n.dx = \frac{m!.n!}{(m+n+1)!}$

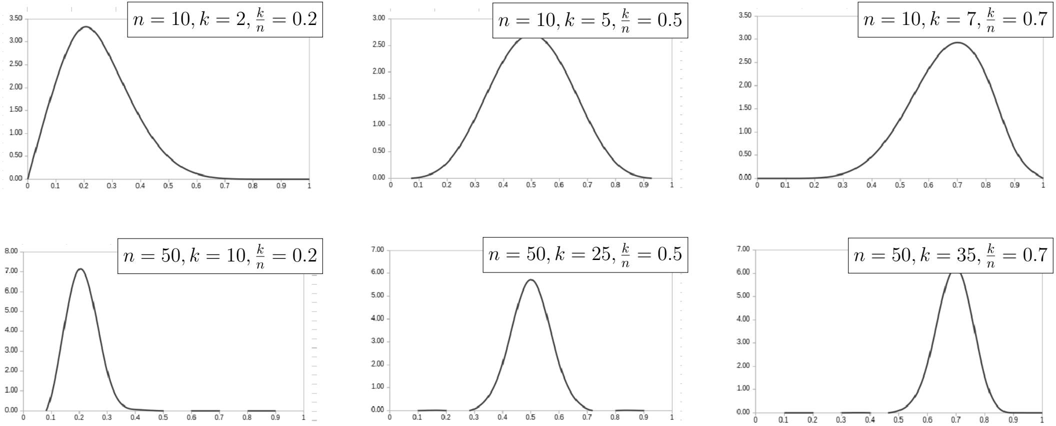
Discussions

- $p_0(\theta) = 1$ [Uniform distribution]
- $D = \langle x_1, x_2, x_1, x_1, \ldots \rangle$ [k x_1s and $(n-k)$ x_2s in any sequence]
- $p_n(\theta) = p(\theta \mid p_0(\theta), D) = \frac{(n+1)!}{k!.(n-k)!}.\theta^k.(1-\theta)^{n-k}$
- $\hat{\theta}_n = \frac{k+1}{n+2}$

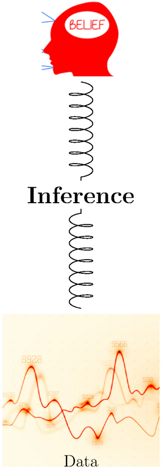
- We get a pdf for θ, rather than a single value
  - More informative, spread tells how reliable it is
- Expected value for θ is similar to MLE
  - ... with two additional experiments with outcomes x_1 and x_2 respectively
- Prior belief in Bayesian estimate moderates the extreme observations
  - Let's assume, we have observed 2 bananas, none is red ($n = 2, k = 0$)
    - MLE: $\theta = \frac{k}{n} = 0$,   Bayesian: $\hat{\theta} = \frac{k+1}{n+2} = \frac{1}{4}$
- Bayesian estimation approaches MLE for large n

Dependence of pdf on data

Assumes uniform prior belief $[p_0(\theta) = 1]$



Priors vs. data (observations)



- We have assumed uniform pdf $p(\theta) = 1$ in this example.
- It is possible of assume other priors
- What determines the priors?
  - Theory or explanations
  - Observation in other domains and inductive generalization
- Cognitive bias

Quiz

Quiz 03-05

End of Module 03-05