Statistics for Data Science - 2

Week 10 Notes

1. **Parameter estimation:** Let $X_1, \ldots, X_n \sim$ iid X, parameter Θ
   Prior distribution of Θ: $\Theta \sim f_\Theta(\theta)$
   Samples: $x_1, \ldots, x_n$, notation $S = (X_1 = x_1, \ldots X_n = x_n)$
   Bayes' rule: posterior $\propto$ likelihood $\times$ prior

   $$P(\Theta = \theta \mid S) = P(S \mid \Theta = \theta) f_\Theta(\theta) / P(S)$$

   In case of discrete: $P(S) = \sum\limits_{\theta} P(S \mid \Theta = \theta) f_\Theta(\theta)$
   In case of continuous: $P(S) = \int\limits_{\theta} P(S \mid \Theta = \theta) f_\Theta(\theta)\ d\theta$
   Posterior mode: $\hat{\theta} = \arg\max_\theta P(S \mid \Theta = \theta) f_\Theta(\theta)$
   Posterior mean: $\mathrm{E}[\Theta \mid S]$, mean of posterior distribution.

2. **Bernoulli(p) samples with uniform prior:** $X_1, \ldots, X_n \sim$ iid Bernoulli($\mathbf{p}$)
   Prior $\mathbf{p} \sim$ Uniform[0, 1]
   Samples: $x_1, \ldots, x_n$
   Posterior: $\mathbf{p} |\ (X_1 = x_1, \ldots X_n = x_n)$
   Posterior density $\propto P(X_1 = x_1, \ldots X_n = x_n \mid \mathbf{p} = p) \times f_{\mathbf{p}}(p)$
   Posterior density $\propto p^w (1-p)^{n-w}$
   $\Rightarrow$ Posterior density: Beta$(w+1, n-w+1)$
   Posterior mean: $\hat{p} = \dfrac{X_1 + X_2 + \ldots + X_n + 1}{n+2}$

3. **Bernoulli(p) samples with beta prior:** $X_1, \ldots, X_n \sim$ iid Bernoulli($\mathbf{p}$)
   Prior $\mathbf{p} \sim$ Beta(α, β)
   $\Rightarrow f_{\mathbf{p}}(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$
   Samples: $x_1, \ldots, x_n$
   Posterior: $\mathbf{p} |\ (X_1 = x_1, \ldots X_n = x_n)$
   Posterior density $\propto P(X_1 = x_1, \ldots X_n = x_n \mid \mathbf{p} = p) \times f_{\mathbf{p}}(p)$
   Posterior density $\propto p^{w+\alpha-1}(1-p)^{n-w+\beta-1}$

   $\Rightarrow$ Posterior density: Beta$(w+\alpha, n-w+\beta)$
   Posterior mean: $\hat{p} = \dfrac{X_1 + X_2 + \ldots + X_n + \alpha}{n+\alpha+\beta}$

4. **Normal samples with unknown mean and known variance:** $X_1, \ldots, X_n \sim$ iid Normal(M, σ^2)

Prior M $\sim$ Normal(μ_0, σ_0^2)
$\Rightarrow f_M(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$
Samples: $x_1, \ldots, x_n$, Sample mean: $\overline{x} = (x_1 + \ldots + x_n)/n$
Posterior: M$|$ $(X_1 = x_1, \ldots X_n = x_n)$
Posterior density $\propto f(X_1 = x_1, \ldots X_n = x_n \mid M = \mu) \times f_M(\mu)$
Posterior density $\propto \exp(-\frac{(x_1-\mu)^2+\ldots+(x_n-\mu)^2}{2\sigma_0^2})\exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$
$\Rightarrow$ Posterior density: Normal
Posterior mean: $\hat{\mu} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}\dfrac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} + \mu_0\dfrac{\sigma^2}{n\sigma_0^2 + \sigma^2}$

5. **Geometric(p) samples with Uniform[0, 1] prior:** $X_1, \ldots, X_n \sim$ iid Geometric($\mathbf{p}$)
Prior $\mathbf{p} \sim$ Uniform[0, 1]
Samples: $x_1, \ldots, x_n$
Posterior: $\mathbf{p}|$ $(X_1 = x_1, \ldots X_n = x_n)$
Posterior density $\propto P(X_1 = x_1, \ldots X_n = x_n \mid \mathbf{p} = p) \times f_{\mathbf{p}}(p)$
Posterior density $\propto p^n(1-p)^{x_1+\ldots+x_n-n}$

$\Rightarrow$ Posterior density: Beta$(n+1, x_1 + \ldots + x_n - n + 1)$
Posterior mean: $\hat{p} = \dfrac{n+1}{X_1 + \ldots + X_n + 2}$

6. **Poisson(λ) samples with gamma prior:** $X_1, \ldots, X_n \sim$ iid Poisson(Λ)
Prior $\Lambda \sim$ Gamma(α, β)
$\Rightarrow f_\Lambda(\lambda) \propto \lambda^{\alpha-1}e^{-\beta\lambda}$
Samples: $x_1, \ldots, x_n$
Posterior: $\Lambda \mid (X_1 = x_1, \ldots X_n = x_n)$
Posterior density $\propto P(X_1 = x_1, \ldots X_n = x_n \mid \Lambda = \lambda) \times f_\Lambda(\lambda)$
Posterior density $\propto e^{-n\lambda}\lambda^{x_1+\ldots+x_n}\lambda^{\alpha-1}e^{-\beta\lambda}$

$\Rightarrow$ Posterior density: Gamma$(x_1 + \ldots + x_n + \alpha, \beta + n)$

Posterior mean: $\hat{\lambda} = \dfrac{X_1 + X_2 + \ldots + X_n + \alpha}{n + \beta}$