STA 601 Homework 5

Lingyun Shao

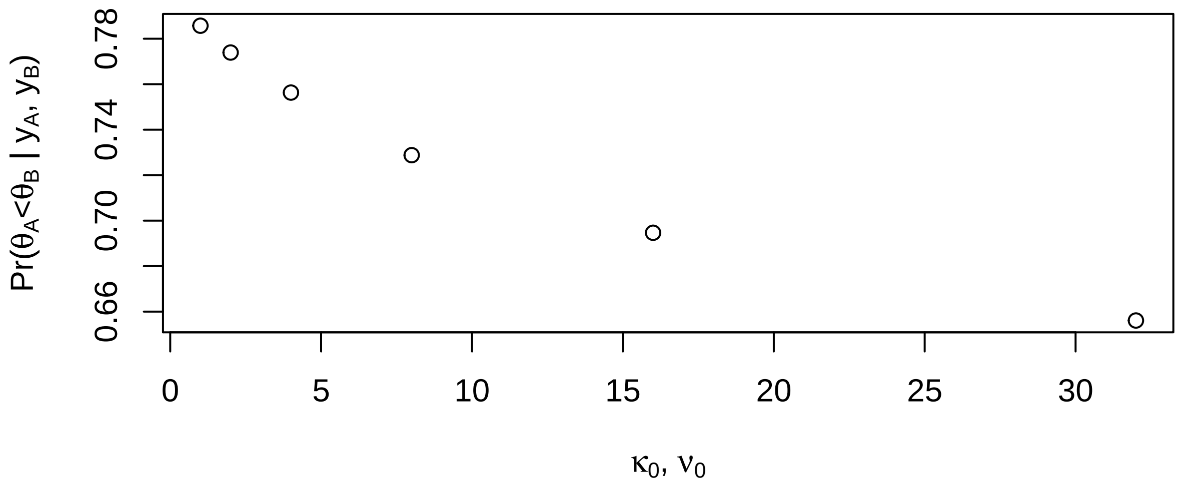
Oct. 09, 2018

5.2

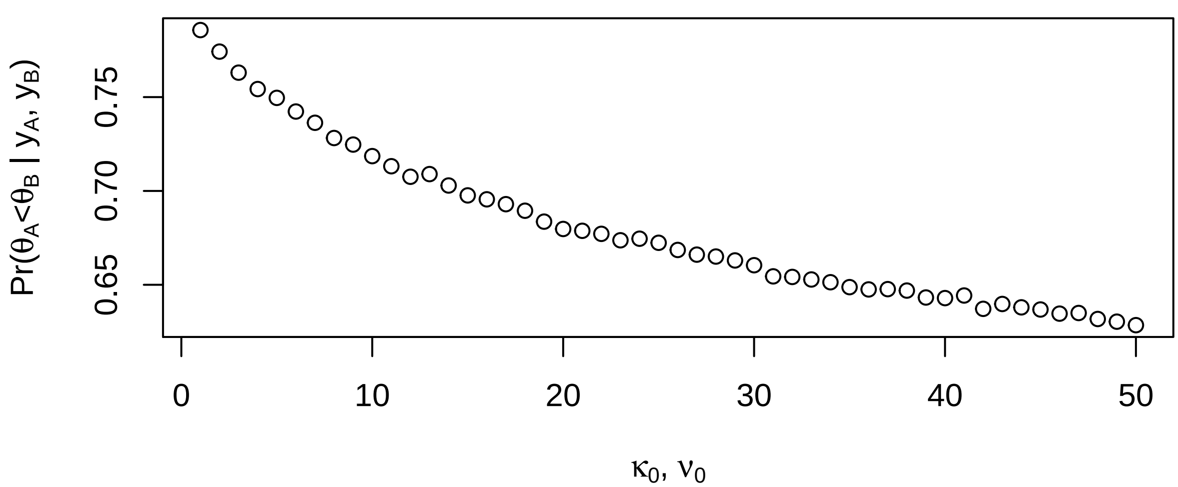
Sensitivity analysis: Thirty-two students in a science classroom were randomly assigned to one of two study methods, A and B, so that $n_A = n_B = 16$ students were assigned to each method. After several weeks of study, students were examined on the course material with an exam designed to give an average score of 75 with a standard deviation of 10. The scores for the two groups are summarized by $\{\bar{y}_A = 75.2, s_A = 7.3\}$ and $\{\bar{y}_B = 77.5, s_B = 8.1\}$. Consider independent, conjugate normal prior distributions for each of θ_A and θ_B, with $\mu_0 = 75$ and $\sigma_0^2 = 100$ for both groups. For each $(\kappa_0, \nu_0) \in \{(1,1),(2,2),(4,4),(8,8),(16,16),(32,32)\}$ (or more values), obtain $Pr(\theta_A < \theta_B | y_A, y_B)$ via Monte Carlo sampling. Plot this probability as a function of $(\kappa_0 = \nu_0)$. Describe how you might use this plot to convey the evidence that $\theta_A < \theta_B$ to people of a variety of prior opinions.

```
est.mc = function(k0, v0 = k0) {
  n.a = n.b = 16
  ybar.a = 75.2
  ybar.b = 77.5
  s.a = 7.3
  s.b = 8.1
  mu0 = 75
  sig20 = 100
  kn.a = k0 + n.a
  kn.b = k0 + n.b
  mun.a = (k0 * mu0 + n.a * ybar.a)/kn.a
  mun.b = (k0 * mu0 + n.b * ybar.b)/kn.b
  vn.a = v0 + n.a
  vn.b = v0 + n.b
  sig2n.a = (v0 * sig20 + (n.a - 1) * s.a^2 + k0 * n.a/kn.a * (ybar.a - mu0)^2)/vn.a
  sig2n.b = (v0 * sig20 + (n.b - 1) * s.b^2 + k0 * n.b/kn.b * (ybar.b - mu0)^2)/vn.b

  nsamp = 100000
  est = NULL
  for(i in 1:length(k0)) {
    sig2.inv.a = rgamma(nsamp, vn.a[i]/2, vn.a[i] * sig2n.a[i]/2)
    theta.a = rnorm(nsamp, mun.a[i], sqrt(1/sig2.inv.a/kn.a[i]))
    sig2.inv.b = rgamma(nsamp, vn.b[i]/2, vn.b[i] * sig2n.b[i]/2)
    theta.b = rnorm(nsamp, mun.b[i], sqrt(1/sig2.inv.b/kn.b[i]))
    est = c(est, mean(theta.a<theta.b))
  }
  return(est)
}
k0 = v0 = 2^(0:5)
plot(k0, est.mc(k0),
     xlab = expression(paste(kappa[0], ", ", nu[0], sep = "")),
     ylab = expression(paste("Pr(", theta[A], "<", theta[B],
                             " | ", y[A], ", ", y[B], ")"), sep = ""))
```




```r
k0 = v0 = 1:50
plot(k0, est.mc(k0),
     xlab = expression(paste(kappa[0], ", ", nu[0], sep = "")),
     ylab = expression(paste("Pr(", theta[A], "<", theta[B],
                             " | ", y[A], ", ", y[B], ")"), sep = ""))
```



I first let $\kappa_0 = \nu_o = 1, 2, 4, 8, 16, 32$ as given in the problem and then added more values, $\kappa_0 = \nu_0 = 1, ..., 50$. By using the function I defined, `est.mc`, I obtained estimated $Pr(\theta_A < \theta_B | y_A, y_B)$ for each pair of κ_0, ν_0. The results of Monte Carlo sampling are displayed above.

From these two plots, we can find that $Pr(\theta_A < \theta_B | y_A, y_B)$ is not very sensitive to different $\kappa_0 = \nu_0$. As is shown in the plot, as κ_0, ν_0 increase, $Pr(\theta_A < \theta_B | y_A, y_B)$ decreases slowly and even when $\kappa_0 = \nu_0 = 50$, the probability is still greater than 0.5.

Notice that the prior belief here is that $\theta_A = \theta_B$ and $\sigma_A^2 = \sigma_B^2$. As our prior degree of belief κ_0, ν_0 increase, the posterior belief will certainly tend to approach the prior belief $\theta_A = \theta_B$, i.e. $Pr(\theta_A < \theta_B | y_A, y_B) = 0.5$. For small $\kappa_0 = \nu_0$ where the data are dominating, we believe $\theta_A < \theta_B$ in certainty. We can obviously see that even with a strong prior belief of $\theta_A = \theta_B$ when $\kappa_0 = \nu_0 = 50$, someone would still tend to believe that $\theta_A < \theta_B$ after he/she observed the data, which is the evidence that $\theta_A < \theta_B$ for people of a variety of prior opinions.

Practice Problem 1

Let $Y_1, ..., Y_n | \theta$ be from i.i.d. $Poisson(\theta)$.

(1) Find the conjugate family of priors.

$$p(\theta | y_1, .., y_n) \propto p(\theta) \prod_{i=1}^{n} p(y_i | \theta) \propto p(\theta) e^{-n\theta} \theta^{\sum_{i=1}^{n} y_i}$$

Judging from the kernel, we know the conjugate family of priors would be gamma distribution.

(2) Find the posterior.

Suppose the prior is $gamma(a, b)$, then we have the kernel of posterior as

$$p(\theta | y_1, .., y_n) \propto p(\theta) \prod_{i=1}^{n} p(y_i | \theta) \propto \theta^{a-1} e^{-b\theta} e^{-n\theta} \theta^{\sum_{i=1}^{n} y_i} = \theta^{a + \sum_{i=1}^{n} y_i - 1} e^{-(b+n)\theta}$$

Thus we know the posterior is $gamma(a + \sum_{i=1}^{n} y_i, b + n)$

(3) Write down the posterior mean as a weighted average.

The posterior mean is

$$\frac{a + \sum_{i=1}^{n} y_i}{b + n} = \frac{a}{b} \frac{b}{b + n} + \frac{\sum_{i=1}^{n} y_i}{n} \frac{n}{b + n} = \frac{a}{b} * \ w_{prior} + \bar{y} * \ w_{data}$$

Practice Problem 2

Let $Y_1, ..., Y_n | \alpha, \beta$ be from i.i.d. $Gamma(\alpha, \beta)$ with α known.

(1) Find the conjugate conjugate family of priors for β.

$$p(\beta | y_1, .., y_n, \alpha) \propto p(\beta) \prod_{i=1}^{n} p(y_i | \alpha, \beta) \propto p(\beta) e^{-\beta \sum_{i=1}^{n} y_i} \beta^{n\alpha}$$

Judging from the kernel, we know the conjugate family of priors for parameter β would be gamma distribution.

(2) Find the posterior.

Suppose the prior is $gamma(a_0, b_0)$, then we have the kernel of posterior as

$$p(\beta | y_1, .., y_n, \alpha) \propto p(\beta) \prod_{i=1}^{n} p(y_i | \alpha, \beta) \propto \beta^{a_0 - 1} e^{-b_0 \beta} e^{-\beta \sum_{i=1}^{n} y_i} \beta^{n\alpha} = \beta^{a_0 + n\alpha - 1} e^{-(b_0 + \sum_{i=1}^{n} y_i)\beta}$$

Thus we know the posterior is $gamma(a_0 + n\alpha, b_0 + \sum_{i=1}^n y_i)$

(3) Give an interpretation of the prior parameters as things like "prior mean", "prior variance", "prior sample size", etc

$$E[\beta|y_1, ..., y_n, \alpha] = \frac{a_0 + n\alpha}{b_0 + \sum_{i=1}^n y_i} = \frac{a_0}{b_0}\frac{b_0}{b_0 + \sum_{i=1}^n y_i} + \frac{n\alpha}{\sum_{i=1}^n y_i}\frac{\sum_{i=1}^n y_i}{b_0 + \sum_{i=1}^n y_i}$$

We notice $\hat{\beta}_{MLE} = \frac{n\alpha}{\sum_{i=1}^n y_i}$. So the posterior mean can be separated as $E[\beta|y_1, ..., y_n, \alpha] = \beta_0 * w_{prior} + \hat{\beta}_{MLE} * w_{data}$, where $\beta_0 = \frac{a_0}{b_0}$ is our prior guess of β, $w_{prior} = \frac{b_0}{b_0 + \sum_{i=1}^n y_i}$, $w_{data} = \frac{\sum_{i=1}^n y_i}{b_0 + \sum_{i=1}^n y_i}$. The hyperparameter b_0 can be viewed as the mean of results for a sequence of prior experiments (like $\sum y_i$ in data).

Practice Problem 3

Let $Y_1, ..., Y_n|\theta$ be from i.i.d. $Bernoulli(\theta)$.

(1) Find the conjugate family of priors

$$p(\theta|y_1, .., y_n) \propto p(\theta)\prod_{i=1}^n p(y_i|\theta) \propto p(\theta)\theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i}$$

Judging from the kernel, we know the conjugate family of priors would be beta distribution.

(2) Find the posterior

Suppose the prior is $beta(a, b)$, then we have the kernel of posterior as

$$p(\theta|y_1, .., y_n) \propto p(\theta)\prod_{i=1}^n p(y_i|\theta) \propto \theta^{a-1}(1-\theta)^{b-1}\theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} = \theta^{a+\sum_{i=1}^n y_i - 1}(1-\theta)^{b+n-\sum_{i=1}^n y_i - 1}$$

Thus we know the posterior is $beta(a + \sum_{i=1}^n y_i, b + n - \sum_{i=1}^n y_i)$.

(3) Give an interpretation of the prior parameters

$$E[\theta|y_1, ..., y_n] = \frac{a + \sum_{i=1}^n y_i}{a + b + n} = \frac{a}{a+b}\frac{a+b}{a+b+n} + \frac{\sum_{i=1}^n y_i}{n}\frac{n}{a+b+n}$$

The posterior mean can be separated as the weighted sum of prior guess of θ, $\frac{a}{a+b}$ and the data mean $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. The prior parameter a can be viewed as the "prior number of success", b can be viewed as the "prior number of failures", and $a + b$ can be viewed as the "prior number of total trials".

(4) Find the posterior predictive distribution.

$$
\begin{aligned}
p(\tilde{y}|y_1,...,y_n) &= \int_\Theta p(\tilde{y}|\theta, y_1,...,y_n)p(\theta|y_1,...,y_n)d\theta \\
&= \int_\Theta p(\tilde{y}|\theta)p(\theta|y_1,...,y_n)d\theta \\
&= \int_\Theta \theta^{\tilde{y}}(1-\theta)^{1-\tilde{y}} B(a+\sum_{i=1}^n y_i, b+n-\sum_{i=1}^n y_i)\theta^{a+\sum_{i=1}^n y_i - 1}(1-\theta)^{b+n-\sum_{i=1}^n y_i - 1} d\theta \\
&= B(a+\sum_{i=1}^n y_i, b+n-\sum_{i=1}^n y_i)\int_\Theta \theta^{\tilde{y}+a+\sum_{i=1}^n y_i - 1}(1-\theta)^{b+n+1-\sum_{i=1}^n y_i - \tilde{y} - 1} d\theta \\
&= \frac{B(a+\sum_{i=1}^n y_i, b+n-\sum_{i=1}^n y_i)}{B(\tilde{y}+a+\sum_{i=1}^n y_i, b+n+1-\sum_{i=1}^n y_i - \tilde{y})} \\
&= \frac{\Gamma(a+b+n)}{\Gamma(a+\sum_{i=1}^n y_i)\Gamma(b+n-\sum_{i=1}^n y_i)} \frac{\Gamma(\tilde{y}+a+\sum_{i=1}^n y_i)\Gamma(b+n+1-\sum_{i=1}^n y_i - \tilde{y})}{\Gamma(a+b+n+1)}
\end{aligned}
$$

Notice that when a, b are both integers, the predictive distribution can be simplified as

$$
p(\tilde{y}|y_1,...,y_n) = \begin{cases} \frac{a+\sum_{i=1}^n y_i}{a+b+n} & \tilde{y} = 1 \\ \frac{b+n-\sum_{i=1}^n y_i}{a+b+n} & \tilde{y} = 0 \end{cases}
$$

Practice Problem 4

Let $Y_1, ..., Y_n|\theta, \sigma^2$ be from i.i.d. $N(\theta, \sigma^2)$ with σ^2 known.

(1) Find the conjugate family of priors for θ.

The posterior distribution of θ with σ^2 known is

$$
p(\theta|y_1,...,y_n,\sigma^2) \propto p(\theta|\sigma^2)p(y_1,...,y_n|\theta,\sigma^2) \propto p(\theta|\sigma^2)e^{-\frac{\sum_{i=1}^n (y_i-\theta)^2}{2\sigma^2}}
$$

Judging from the kernel, the conjugate prior should have a form like $e^{c_1(\theta-c_2)^2}$, so a simple choice of conjugate family of priors would be normal distribution.

(2) Find the posterior.

Suppose the prior is $Normal(\mu_0, \tau_0^2)$, then we have the kernel of posterior as

$$
p(\theta|y_1,...,y_n,\sigma^2) \propto p(\theta|\sigma^2)p(y_1,...,y_n|\theta,\sigma^2) \propto e^{-\frac{(\theta-\mu_0)^2}{2\tau_0^2}} e^{-\frac{\sum_{i=1}^n (y_i-\theta)^2}{2\sigma^2}}
$$

Thus we know the posterior is $beta(a+\sum_{i=1}^n y_i, b+n-\sum_{i=1}^n y_i)$.

Adding the terms in the exponents and ignoring the -1/2 for the moment, we have

$$
\frac{1}{\tau_0^2}(\theta^2 - 2\theta\mu_0 + \mu_0^2) + \frac{1}{\sigma^2}(\sum y_i^2 - 2\theta \sum y_i + n\theta^2) = a\theta^2 - 2b\theta + c
$$

where $a = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$, $b = \frac{\mu_0}{\tau_0^2} + \frac{\sum y_i}{\sigma^2}$ and c is some constant. So rearranging the terms, we have

$$
p(\theta|y_1,...,y_n,\sigma^2) \propto e^{-\frac{(\theta-b/a)^2}{2a}}
$$

Thus we know from the kernel form that the posterior is $N(\mu_n, \tau_n^2)$, where $\tau_n^2 = 1/a = \frac{1}{1/\tau_0^2 + n/\sigma^2}$, $\mu_n = b/a = (\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2})\tau_n^2$.

(3) Give an interpretation of the prior parameters

Inverse variance is often referred to as the precision. For normal model, denote the precision as below

$$\tilde{\sigma}^2 = 1/\sigma^2 = sampling \;\; precision, \;\; \tilde{\tau}_0^2 = 1/\tau_0^2 = prior \;\; precision, \;\; \tilde{\tau}_n^2 = 1/\tau_n^2 = posterior \;\; precision$$

So the posterior mean can be written as

$$\mu_n = \frac{\tilde{\tau}_0^2}{\tilde{\tau}_0^2 + n\tilde{\sigma}^2}\mu_o + \frac{n\tilde{\sigma}^2}{\tilde{\tau}_0^2 + n\tilde{\sigma}^2}\bar{y}$$

We can find that the posterior mean is a weighted average of the prior mean and the sample mean. The weight on the sample mean is n/σ^2, the sampling precision of the sample mean. The weight on the prior mean is $1/\tau_0^2$, the prior precision. If the prior mean were based on κ_0 prior observations from the same population as $Y_1, ..., Y_n$, then we might want to set $\tau_0^2 = \sigma^2/\kappa_0$. Thus, the posterior mean reduces to

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y}$$

(4) Find the posterior predictive distribution.

$\tilde{y}$ can be separated as $\tilde{y} = \theta + \sigma Z$, where $Z \sim N(0, 1)$ and θ and Z are independent. Using the fact that the sum of independent normal random variables is also normal, we know $\tilde{y}$ should follow a normal distribution.

$$\begin{aligned} E[\tilde{y}|y_1, ..., y_n, \sigma^2] &= E[\theta + \sigma Z|y_1, ..., y_n, \sigma^2] \\ &= E[\theta|y_1, ..., y_n, \sigma^2] + E[\sigma Z|y_1, ..., y_n, \sigma^2] \\ &= \mu_n + 0 \times \sigma \\ &= \mu_n \end{aligned}$$

$$\begin{aligned} Var[\tilde{y}|y_1, ..., y_n, \sigma^2] &= Var[\theta + \sigma Z|y_1, ..., y_n, \sigma^2] \\ &= Var[\theta|y_1, ..., y_n, \sigma^2] + Var[\sigma Z|y_1, ..., y_n, \sigma^2] \\ &= \tau_n^2 + 1 \times \sigma^2 \\ &= \tau_n^2 + \sigma^2 \end{aligned}$$

Therefore, the predictive distribution of $\tilde{y}$ is $N(\mu_n, \tau_n^2 + \sigma^2)$ where τ_n^2 and μ_n are given as in (2).

Quiz 1

1.

(a) State Bayes' theorem.

Suppose we have some prior distribution $p(\theta)$ of parameter θ, a sampling model $p(y|\theta)$, then the posterior of θ is

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int_\Theta p(\theta)p(y|\theta)d\theta}$$

(b) Define finite exchangeability.

Let $p(y_1, ..., y_n)$ be joint density of $Y_1, ..., Y_n$. If $p(y_1, ..., y_n) = p(y_{\pi_1}, ..., y_{\pi_n})$ for all permutations π of $\{1, .., n\}$, then $Y_1, .., Y_n$ are exchangeable.

(c) State the Fisher factorization theorem.

If $p(y|\theta)$ is a p.d.f. then T is sufficient for θ if and only if $\exists$ non-negative functions g, h s.t. $p(y|\theta) = h(y)g(T(y)|\theta)$.

(d) TRUE or FALSE: An interval with 95% Bayesian coverage describes your information about the location of the true value of θ after observing $Y = y$.

TRUE

(e) TRUE or FALSE: The posterior predictive distribution belongs to the same family as the sampling distribution.

FALSE

2.

The Weibull distribution with unknown scale parameter θ is given by the following parametrization $p(y|\theta) = \frac{\beta}{\theta} y^{\beta-1} \exp(-\frac{y^\beta}{\theta})$ where β is a known shape parameter.

(a) Find the MLE for θ.

Suppose we have i.i.d. $Y_1, ..., Y_n \sim Weibull(\theta)$ with β given. Then the likelihood function can be written as

$$\mathcal{L}(\theta|y_1, ..., y_n) = \prod_{i=1}^{n} p(y_i|\theta) = (\frac{\beta}{\theta})^n (\prod y_i)^{\beta-1} \exp(-\frac{\sum y_i^\beta}{\theta})$$

The log-likelihood $\log \mathcal{L} = -n\log(\theta) - \frac{\sum y_i^\beta}{\theta} + c$. Set $\frac{\partial \mathcal{L}}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum y_i^\beta}{\theta^2} = 0$, we have $\hat{\theta}_{MLE} = \frac{\sum y_i^\beta}{n}$.

(b) Derive the conjugate family of priors for this sampling model.

Let $\lambda = 1/\theta$, we can find the conjugate family of priors of λ for this sampling models.

$$p(\lambda|y_1, ..., y_n) \propto p(\lambda) \prod p(y_i|\lambda) \propto p(\lambda)\lambda^n \exp(-\sum y_i^\beta \lambda)$$

Judging from the kernel, we know the prior should include terms like $\lambda^{c_1} \exp(-c_2\lambda)$. The simple choice of conjugate family of priors for λ would be gamma distribution.

(c) Write the posterior mean as a weighted sum of the prior mean and the MLE. Provide an interpretation to the parameters in the prior.

Suppose our prior for λ is $gamma(a, b)$, the kernel of posterior can be written as

$$p(\lambda|y_1, ..., y_n) \propto p(\lambda) \prod p(y_i|\lambda) \propto \lambda^{a-1} \exp(-b\lambda)\lambda^n \exp(-\sum y_i^\beta \lambda) = \lambda^{a+n-1} \exp(-(b + \sum y_i^\beta)\lambda)$$

From the form of the posterior kernel, we know $\lambda|y \sim gamma(a + n, b + \sum |y_i|)$.

$E[\lambda|y_1, ..., y_n] = \frac{a+n}{b+\sum |y_i|} = \frac{a}{b} \frac{b}{b+\sum |y_i|} + \frac{n}{\sum |y_i|} \frac{\sum |y_i|}{b+\sum |y_i|} = \lambda_0 * w_{prior} + \hat{\lambda}_{MLE} * w_{data}$, where $\lambda_0 = \frac{a}{b}$ is our prior guess of $\lambda = 1/\theta$ and $\hat{\lambda}_{MLE} = 1/\hat{\theta}_{MLE}$ is derived from the data. Prior parameter b can be viewed as the degree of trust in our prior belief.

3.

Let X be an Exponential distribution with mean λ. Your goal is to study the mean without injecting too much prior information.

(a) Derive Jeffrey's prior for λ and state if it is a proper distribution. $p(x|\lambda) = \frac{1}{\lambda}\exp(-\frac{x}{\lambda})$. So we have its Fisher information

$$\begin{aligned} I(\lambda) &= -E[\frac{\partial^2 \log p}{(\partial\lambda)^2}] \\ &= -E[\frac{1}{\lambda^2} - \frac{2x}{\lambda^3}] \\ &= -\frac{1}{\lambda^2} + \frac{2\lambda}{\lambda^3} \\ &= \frac{1}{\lambda^2} \end{aligned}$$

Jeffrey's prior for λ $p(\lambda) \propto \sqrt{I(\lambda)} = 1/\lambda$. Since the support of λ is $\mathbb{R}^+$ and $\int_{\mathbb{R}^+}(1/\lambda)d\lambda = \infty$, so it's not a proper distribution.

(b) Derive the posterior and state if it is proper.

$$p(\lambda|x) \propto p(\lambda)p(x|\lambda) = \frac{1}{\lambda}\frac{1}{\lambda}\exp(-\frac{x}{\lambda}) = \frac{1}{\lambda^2}\exp(-\frac{x}{\lambda})$$

From the form of the kernel, we know $\lambda|x \sim \textit{Inverse Gamma}(1, x)$, so it's proper.

Quiz 2

1.

(a) State Bayes' theorem. (duplicate)

(b) Define independent random variables.

$Y_1, ..., Y_n$ are independent random variables if for every collection of n sets $\{A_1, ..., A_n\}$ we have

$$Pr(Y_1 \in A_1, ..., Y_n \in A_n) = Pr(Y_1 \in A_1) \times \cdots \times Pr(Y_n \in A_n)$$

(c) Define the notion of conjugate families.

A class $\mathcal{P}$ of prior distributions for θ is called conjugate for a sampling model $p(y|\theta)$ if

$$p(\theta) \in \mathcal{P} \Rightarrow p(\theta|y) \in \mathcal{P}$$

(d) TRUE or FALSE: A frequentist interval with 95% coverage is a random interval with no post-experimental interpretation.

TRUE

(e) TRUE or FALSE: Predictive distribution does not depend on any unknown quantities.

TRUE

2.

Extreme events are frequently described by the double exponential distribution. Consider the parametrization in this case $p(y|\theta) = \frac{1}{2\theta}\exp(-\frac{|y|}{\theta})$

(a) Find the MLE for θ.

$$\mathcal{L}(\theta|y) = \prod_{i=1}^{n} p(y_i|\theta) = (\frac{1}{2\theta})^n \exp(-\frac{\sum |y_i|}{\theta})$$

$\log \mathcal{L} = -n\log\theta - \frac{\sum|y_i|}{\theta} + c$ Set $\frac{\partial \log \mathcal{L}}{\partial\theta} = -\frac{n}{\theta} + \frac{\sum|y_i|}{\theta^2} = 0$, we have $\hat{\theta}_{MLE} = \frac{\sum|y_i|}{n}$

(b) Derive the conjugate family of priors for this sampling model.

Let $\lambda = 1/\theta$, we can find the conjugate family of priors of λ for this sampling models.

$$p(\lambda|y_1,...,y_n) \propto p(\lambda)\prod p(y_i|\lambda) \propto p(\lambda)\lambda^n \exp(-\sum|y_i|\lambda)$$

Judging from the kernel, we know the prior should include terms like $\lambda^{c_1}\exp(-c_2\lambda)$. The simple choice of conjugate family of priors for λ would be gamma distribution.

(c) Write the posterior mean as a weighted sum of the prior mean and the MLE. Provide an interpretation to the parameters in the prior.

Suppose our prior for λ is $gamma(a, b)$, the kernel of posterior can be written as

$$p(\lambda|y_1,...,y_n) \propto p(\lambda)\prod p(y_i|\lambda) \propto \lambda^{a-1}\exp(-b\lambda)\lambda^n \exp(-\sum|y_i|\lambda) = \lambda^{a+n-1}\exp(-(b+\sum|y_i|)\lambda)$$

From the form of the posterior kernel, we know $\lambda|y \sim gamma(a+n, b+\sum y_i^\beta)$.

$E[\lambda|y_1,...,y_n] = \frac{a+n}{b+\sum y_i^\beta} = \frac{a}{b}\frac{b}{b+\sum y_i^\beta} + \frac{n}{\sum y_i^\beta}\frac{\sum y_i^\beta}{b+\sum y_i^\beta} = \lambda_0 * w_{prior} + \hat{\lambda}_{MLE} * w_{data}$, where $\lambda_0 = \frac{a}{b}$ is our prior guess of $\lambda = 1/\theta$ and $\hat{\lambda}_{MLE} = 1/\hat{\theta}_{MLE}$ is derived from the data. Prior parameter b can be viewed as the degree of trust in our prior belief.

3.

Let X be a Poisson distribution with mean θ. Your goal is to study the mean without injecting too much prior information.

(a) Derive Jeffrey's prior for μ and state if it is a proper distribution.

$p(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}$. So we have its Fisher information

$$\begin{aligned} I(\theta) &= -E[\frac{\partial^2 \log p}{(\partial\theta)^2}] \\ &= -E[-\frac{x}{\theta^2}] \\ &= \frac{1}{\theta} \end{aligned}$$

Jeffrey's prior for θ, $p(\theta) \propto \sqrt{I(\theta)} = 1/\sqrt{\theta}$. Since the support of θ is $[0,\infty)$ and $\int_{[0,\infty)}(1/\sqrt{\theta})d\lambda = \infty$, so it's not a proper distribution.

(b) Derive the posterior and state if it is proper.

$$p(\theta|x) \propto p(\theta)p(x|\theta) = \frac{1}{\sqrt{\theta}} \frac{e^{-\theta}\theta^x}{x!} = \frac{e^{-\theta}\theta^{x-1/2}}{x!}$$

From the form of the kernel, we know $\theta|x \sim \ Gamma(x+1/2, 1)$, so it's proper.

Quiz 3

1.

(a) State Bayes' theorem (duplicate)

(b) Mathematically define the median of a distribution.

Median m is the quantile in a distribution of random variable X s.t

$$Pr(X \leq m) = Pr(X \geq m)$$

(c) Define posterior predictive distribution.

The posterior predictive distribution is, given prior $p(\theta)$ data $y_1, ..., y_n$, the probability distribution to predict a new sample $\tilde{y}$.

$$p(\tilde{y}|y_1, .., y_n) = \int_{\Theta} p(\tilde{y}|\theta, y_1, ..., y_n)p(\theta|y_1, ..., y_n)d\theta$$

(d) TRUE or FALSE: Highest posterior density credible regions must be intervals.

FALSE

(e) TRUE or FALSE: All sequences of independent random variables are exchangeable

TRUE

2.

The Maxwell distribution describes particle speeds in idealized gases. Consider the following parametrization in this case $p(y|\theta) = (2/\pi)^{1/2}\theta^{3/2}y^2 \exp(-\frac{\theta y^2}{2})$.

(a) Find the MLE for θ.

$$\mathcal{L}(\theta|y) = \prod_{i=1}^{n} p(y_i|\theta) = (2/\pi)^{n/2}\theta^{3n/2}(\prod y_i^2) \exp(-\frac{\theta \sum y_i^2}{2})$$

$\log \mathcal{L} = 3n/2 \times \log\theta - \frac{\theta}{2}(\sum y_i^2) + c$ Set $\frac{\partial \log \mathcal{L}}{\partial \theta} = \frac{3n}{2\theta} - \frac{\sum y_i^2}{2} = 0$, we have $\hat{\theta}_{MLE} = \frac{3n}{\sum y_i^2}$

(b) Derive the conjugate family of priors for this sampling model.

Say $p(\theta)$ is the prior, $p(\theta|y) \propto p(\theta)p(y|\theta) \propto p(\theta)\theta^{3n/2} \exp(-\frac{\sum y_i^2 \theta}{2})$. Then the conjugate prior must have kernel like $\theta^{c_1} \exp(-c_2\theta)$. The simple choice is $Gamma(a, b)$, whose p.d.f. is $\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}$.

(c) Write the posterior mean as a weighted sum of the prior mean and the MLE. Provide an interpretation to the parameters in the prior.

$$p(\theta|y) \propto \theta^{3n/2+a-1} e^{-(b+\frac{\sum y_i^2}{2})\theta}$$

Judging from the kernel, $\theta|y \sim Gamma(a+3n/2, b+\frac{\sum y_i^2}{2})$

$E(\theta|y) = \frac{3n/2+a}{b+\sum y_i^2/2} = \frac{b}{b+\sum y_i^2/2}\frac{a}{b} + \frac{\sum y_i^2/2}{b+\sum y_i^2/2}\frac{3n/2}{\sum y_i^2/2} = w_{prior} * \mu_0 + w_{data} * \hat{\theta}_{MLE}$.

Where $\mu_0 = a/b$ is our prior guess for θ, b can be viewed as our prior degree of trust in the prior information.

3.

Let X be a Normal distribution with mean μ and variance 1. Your goal is to study the mean without injecting too much prior information.

(a) Derive Jeffrey's prior for μ and state if it is a proper distribution.

$p(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$. So we have its Fisher information

$$\begin{aligned} I(\mu) &= -E[\frac{\partial^2 \log p}{(\partial \mu)^2}] \\ &= -E[-1] \\ &= 1 \end{aligned}$$

Jeffrey's prior for μ, $p(\mu) \propto \sqrt{I(\mu)} = 1$. Since the support of μ is $\mathbb{R}$ and $\int_{\mathbb{R}} 1 d\lambda = \infty$, so it's not a proper distribution.

(b) Derive the posterior and state if it is proper.

$$p(\mu|x) \propto p(\mu)p(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

From the form of the kernel, we know $\mu|x \sim \ N(x, 1)$, so it's proper.