STAT 581
Homework #3
10/21/2013

Solutions for Hoff Exercises 5.1, 5.3, 5.4, 5.8, and LC 4.2.2

5.1 Let $X_1, \ldots, X_n \sim$ iid $N(\theta, \sigma^2)$, where σ^2 is known. Using the prior distribution $\theta \sim N(\theta_0, \sigma^2/\kappa_0)$,
5.1.a Obtain the Bayes estimator under squared error loss.

Under squared error loss, the Bayes estimator will be the posterior mean $\mathrm{E}[\theta|\boldsymbol{X}]$. In this case, looking at

$$\pi(\theta|\boldsymbol{X}) \propto_\theta p(\boldsymbol{X}|\theta)\pi(\theta)$$

and plugging in the known expressions for $p(\boldsymbol{X}|\theta)$ and $\pi(\theta)$, we find that $\pi(\theta|\boldsymbol{X})$ is proportional to the kernel of a normal density with mean

$$\frac{\kappa_0\theta_0 + n\bar{X}}{\kappa_0 + n}$$

and variance

$$\frac{\sigma^2}{\kappa_0 + n}.$$

Therefore, the Bayes estimator is just the posterior mean, namely

$$\delta_\pi = \frac{\kappa_0\theta_0 + n\bar{X}}{\kappa_0 + n} = \frac{\kappa_0}{\kappa_0 + n}\theta_0 + \frac{n}{\kappa_0 + n}\bar{X}.$$

■

5.1.b Compute the risk function of the Bayes estimator.

By definition,

$$R(\theta, \delta_\pi) = \mathrm{E}_{\boldsymbol{X}|\theta}[(\delta_\pi - \theta)^2]$$

Plugging in the expression for δ_π from part (a), we see that this expectation just requires us to compute first and second moments of $\bar{X}$, which is easily done, since $\bar{X}$ is normally distributed. The result of the computations is

$$R(\theta, \delta_\pi) = \frac{\kappa_0^2(\theta - \theta_0)^2 + n\sigma^2}{(n + \kappa_0)^2}.$$

■

5.1.c Suppose the true population mean is μ. For what values of θ_0 and κ_0 will the Bayes estimator beat the sample mean, in terms of risk? Write your answer in terms of $(\mu - \theta_0)^2$, κ_0, σ^2, and n.

In class we showed that the risk of the sample mean is a constant value, σ^2/n. So the question is just for which values we have

$$\frac{\kappa_0^2(\mu - \theta_0)^2 + n\sigma^2}{(n + \kappa_0)^2} < \frac{\sigma^2}{n}.$$

We can simplify this to

$$(\mu - \theta_0)^2 < 2\frac{\sigma^2}{\kappa_0} + \frac{\sigma^2}{n},$$

that is, the Bayes estimator beats the sample mean when μ and θ_0 are "close". ■

5.3 Let $\boldsymbol{X} = X_1, \ldots, X_n \sim$ iid $N(\theta, \sigma^2)$. Consider the prior distribution $\pi(\theta, \sigma^2)$ where $\pi(\theta|\sigma^2)$ is the $N(\theta_0, \sigma^2/\kappa_0)$ density and $\pi(\sigma^2)$ is the inverse-gamma density, so that $1/\sigma^2 \sim \text{Gamma}(a = \nu_0/2, b = \nu_0\sigma_0^2/2)$ and $1/\sigma^2$ has expectation $a/b = 1/\sigma_0^2$. For this model and prior, obtain $\pi(\sigma^2|\theta, \boldsymbol{X})$, $\pi(\theta|\sigma^2, \boldsymbol{X})$, $\pi(\sigma^2|\boldsymbol{X})$, and $\pi(\theta|\boldsymbol{X})$. If you prefer, you can write everything in terms of $\tau^2 = 1/\sigma^2$ instead.

It's easiest to start with

$$\pi(\tau^2|\theta, \boldsymbol{X}) \propto_\tau \pi(\boldsymbol{X}|\tau^2, \theta)\pi(\tau^2|\theta) \propto_\tau \pi(\boldsymbol{X}|\tau^2, \theta)\pi(\theta|\tau^2)\pi(\tau^2).$$

Plugging in these known densities and reducing terms, we see that $\pi(\tau^2|\theta, \boldsymbol{X})$ is proportional to the kernel of a Gamma density with shape parameter $(n + \nu_0 + 1)/2$ and rate parameter $\frac{1}{2}(\nu_0\sigma_0^2 + \kappa_0(\theta - \theta_0^2) + \sum(X_i - \theta)^2)$. That is,

$$\tau^2|\theta, \boldsymbol{X} \sim \text{Gamma}\left(\frac{(n + \nu_0 + 1)}{2}, \frac{1}{2}\left(\nu_0\sigma_0^2 + \kappa_0(\theta - \theta_0^2) + \sum(X_i - \theta)^2\right)\right).$$

We apply a similar procedure for $\pi(\theta|\tau^2, \boldsymbol{X})$.

$$\pi(\theta|\tau^2, \boldsymbol{X}) \propto_\theta \pi(\boldsymbol{X}|\theta, \tau^2)\pi(\theta|\tau^2).$$

We can then plug in these known densities and find that $\pi(\theta, \tau^2, \boldsymbol{X})$ is proportional to the kernel of a normal density. The result is

$$\theta|\tau^2, \boldsymbol{X} \sim N\left(\frac{n\bar{X} + \theta_0\kappa_0}{n + \kappa_0}, \frac{1}{\tau^2(n + \kappa_0)}\right).$$

For $\pi(\tau^2|\boldsymbol{X})$, we note

$$\begin{aligned}\pi(\tau^2|\boldsymbol{X}) &= \frac{\pi(\tau^2,\boldsymbol{X})}{\pi(\boldsymbol{X})} \\ &= \frac{\int_\theta \pi(\boldsymbol{X},\theta,\tau^2)d\theta}{\pi(\boldsymbol{X})} \\ &\propto_{\tau^2} \int_\theta \pi(\boldsymbol{X}|\tau^2,\theta)\pi(\theta|\tau^2)\pi(\tau^2)d\theta\end{aligned}$$

Plugging in densities and computing the integral, we find

$$\tau^2|\boldsymbol{X} \sim \text{Gamma}\left(\frac{n+\nu_0}{2}, \frac{1}{2}\left(\nu_0\sigma_0^2 + \frac{n\kappa_0}{n+\kappa_0}(\theta_0-\bar{X})^2 + \sum(X_i-\bar{X})^2\right)\right).$$

Finally, we have

$$\pi(\theta|\boldsymbol{X}) \propto_\theta \int_{\tau^2} \pi(\boldsymbol{X}|\tau^2,\theta)\pi(\theta|\tau^2)\pi(\tau^2)d\tau^2.$$

Computing this integral and examining the resulting density, we find that $\theta|\boldsymbol{X}$ follows a non-standardized t-distribution with degrees of freedom parameter $\nu = \nu_0 + n$, mean parameter $\mu = \frac{\kappa_0\theta_0+n\bar{X}}{\kappa_0+n}$, and squared scale parameter

$$\sigma^2 = \frac{1}{\nu_0+n}\left(\frac{\nu_0\sigma_0^2+\kappa_0\theta_0^2+\sum X_i^2}{\kappa_0+n} - \left(\frac{\kappa_0\theta_0+n\bar{X}}{\kappa_0+n}\right)^2\right).$$

■

5.4 Let $X_1,\ldots,X_n \sim$ iid Poisson(θ), and let $\theta \sim$ Gamma(a,b) with expectation a/b. Suppose you observe $X_1,\ldots,X_n$ and want to predict the value of a future observation from this population. Obtain the form of the *predictive distribution* $p(X^*|x_1,\ldots,x_n)$ based on the Poisson model $X_1,\ldots,X_n,X^* \sim$ iid Poisson(θ) and the Gamma(a,b) prior distribution. Note that the predictive distribution does not depend on any unknown parameters, meaning that you can actually use it to make predictions. Letting μ be the true population mean, what does the predictive distribution converge to as $n \to \infty$? Explain why this limiting distribution makes sense.

First, we note that the posterior predictive distribution $p(X^*|x_1,\ldots,x_n)$ can be computed as

$$\begin{aligned}p(X^*|x_1,\ldots,x_n) &= \int_\theta p(X^*,\theta|x_1,\ldots,x_n)\,d\theta \\ &= \int_\theta p(X^*|\theta,x_1,\ldots,x_n)p(\theta|x_1,\ldots,x_n)\,d\theta \\ &= \int_\theta p(X^*|\theta)p(\theta|x_1,\ldots,x_n)\,d\theta.\end{aligned}$$

We already know $p(X^*|\theta)$, so we will be able to evaluate this integral once we know the posterior density $p(\theta|x_1,\ldots,x_n)$.

Now,

$$\begin{aligned}
p(\theta|x_1,\ldots,x_n) &= \frac{p(x_1,\ldots,x_n|\theta)\pi(\theta)}{p(x_1,\ldots,x_n)} \\
&\propto_\theta p(x_1,\ldots,x_n|\theta)\pi(\theta) \\
&\propto_\theta \left(\prod_{i=1}^n \frac{\theta_i^x}{x_i} e^{-\theta}\right) \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \\
&\propto_\theta \theta^{n\bar{x}+a-1} e^{-\theta(b+n)},
\end{aligned}$$

which we recognize as the kernel of a Gamma$(n\bar{x}+a, b+n)$ distribution.

Returning now to the posterior predictive distribution, we have

$$\begin{aligned}
p(X^*|x_1,\ldots,x_n) &= \int_\theta p(X^*|\theta)p(\theta|x_1,\ldots,x_n)\,d\theta \\
&= \int_0^\infty \frac{\theta^{X^*}}{X^*!} e^{-\theta} \frac{(b+n)^{n\bar{x}+a}}{\Gamma(n\bar{x}+a)} \theta^{n\bar{x}+a-1} e^{-\theta(b+n)}\,d\theta \\
&= \frac{(b+n)^{n\bar{x}+a}}{X^*!\Gamma(n\bar{x}+a)} \int_0^\infty \theta^{X^*+n\bar{x}+a-1} e^{-\theta(b+n+1)}\,d\theta.
\end{aligned}$$

Now we can either recognize the integrand as the kernel of a Gamma pdf or consult an integral table to obtain the result

$$p(X^*|x_1,\ldots,x_n) = \frac{(b+n)^{n\bar{x}+a}}{X^*!\Gamma(n\bar{x}+a)} \frac{\Gamma(X^*+n\bar{x}+a)}{(b+n+1)^{X^*+n\bar{x}+a}}.$$

With a little work, we can rewrite this expression as

$$\begin{aligned}
v = \frac{1}{X^*!} \frac{\Gamma(X^*+a+n\bar{x})}{\Gamma(a+n\bar{x})(a+n\bar{x})^{X^*}} \left(\frac{a+n\bar{x}}{b+n+1}\right)^{X^*} \\
\times \left(\left(1-\frac{1}{b+n+1}\right)^{b+n+1}\right)^{\bar{x}} \left(1-\frac{1}{b+n+1}\right)^{a-(b+1)\bar{x}}.
\end{aligned}$$

Looking at the limits of each of these terms as n goes to infinity, we find

$$\frac{1}{X^*!} \to \frac{1}{X^*!}$$

$$\frac{\Gamma(X^*+a+n\bar{x})}{\Gamma(a+n\bar{x})(a+n\bar{x})^{X^*}} \to 1$$

$$\left(\frac{a+n\bar{x}}{b+n+1}\right)^{X^*} \to \bar{x}^{X^*}$$

$$\left(\left(1-\frac{1}{b+n+1}\right)^{b+n+1}\right)^{\bar{x}} \to e^{-\bar{x}}$$

and

$$\left(1-\frac{1}{b+n+1}\right)^{a-(b+1)\bar{x}} \to 1.$$

So altogether we have $p(X^*|x_1,\ldots,x_n) \to e^{-\bar{x}}\frac{\bar{x}^{X^*}}{X^*!}$. That is, $X^*|x_1,\ldots,x_n$ in asymptotically Poisson$(\bar{x})$ distributed. Furthermore, by the law of large numbers, we have $\bar{x} \to \mu$, so the limiting predictive distribution is Poisson(μ), i.e. the true distribution of $X^*|\theta$. This makes sense because as the amount of data grows, we expect the data to give us a correct inference about the parameter, and the influence of the prior distribution should shrink to nil as we learn more from the data. ■

5.8 Consider Bayesian inference using a posterior density $\pi(\theta|x)$:
5.8.a Find the form of the Bayes estimator under absolute loss $L(\theta,d) = |\theta - d|$, and prove your result.

Let's start with the assumptions that $\theta|x$ has a continuous distribution and that $\pi(\theta|x)$ is a nice enough function that we can interchange differentiation and integration.

The Bayes estimator will minimize posterior risk $R(\pi,d|x)$. By definition, we have

$$\begin{aligned} R(\pi,d|x) &= \int_\theta |\theta-d|\pi(\theta|x)d\theta \\ &= \int_{\theta<d}(d-\theta)\pi(\theta|x)d\theta + \int_{\theta\geq d}(\theta-d)\pi(\theta|x)d\theta. \end{aligned}$$

To find the minimizing value of d, we differentiate posterior risk with respect to d and set the derivative equal to 0.

$$0 = \frac{\partial}{\partial d}R(\pi,d|x) = \int_{\theta<d}\pi(\theta|x)d\theta - \int_{\theta\geq d}\pi(\theta|x)d\theta.$$

This equation is solve by any d for which $\int_{\theta<d}\pi(\theta|x)d\theta = \int_{\theta\geq d}\pi(\theta|x)d\theta$. That is, d can be any *posterior median*.

If we don't make the assumptions about the distribution of $\theta|x$, it is still possible to show that posterior risk is minimized by a posterior median. One possible method is to proceed directly by fixing a value of x and assuming some estimator d' such that $\pi(\theta \leq d'|x) < 1/2$. It's then possible to show that $R(\pi,d|x) \leq R(\pi,d'|x)$, so that it's not possible for a non-median to have lower posterior risk than a median. ■

5.8.b Find the form of the Bayes estimator under zero-one loss $L(\theta,d) = 1(\theta \neq d)$ for the case that $\pi(\theta|x)$ is discrete.

Again, the Bayes estimator will minimize the posterior risk. In this case, the posterior risk is

$$\begin{aligned} R(\pi, d|x) &= \int_\theta 1(\theta \neq d)\pi(\theta|x)d\theta \\ &= \sum_{\theta \in \Theta} 1(\theta \neq d)\pi(\theta|x) \\ &= 1 - \pi(d|x), \end{aligned}$$

which is minimized if we maximize $\pi(d|x)$. That is, the Bayes estimator is any *posterior mode*. ■

LC 4.2.2 Consider the exaple of sequential binomial sampling (LC Example 4.2.1). Let X be the number of successes in n Bernoulli trials with success probability p.
4.2.2.a Suppose that the number of Bernoulli trials performed is a prespecified number n, so that we have the binomial sampling model

$$P(X = x) = \binom{n}{k} p^x (1-p)^{n-x},\ x = 0, 1, \ldots, n.$$

Calculate the Bayes risk of the Bayes estimator and the UMVU estimator of p.

The algebra for this problem is rather unpleasant, so we just give setup and answers.

We know the Bayes estimator is given by

$$\delta_\pi(X) = \frac{a + X}{a + b + n}$$

and the UMVU estimator is just X/n.

The Bayes risk of the Bayes estimator is

$$\mathrm{E}_p[\mathrm{E}_{X|p}[(\delta_\pi(X) - p)^2]].$$

Performing the inner expectation yields (after a lot of algebra)

$$\mathrm{E}_{X|p}[(\delta_\pi(X) - p)^2] = \frac{(p(a+b+2n) + a)^2 + np(1-p)}{(a+b+n)^2}.$$

Subsequently performing the outer expectation yields

$$\mathrm{E}_p[\mathrm{E}_{X|p}[(\delta_\pi(X) - p)^2]] = \frac{ab}{(a+b)(a+b+1)(a+b+n)}.$$

On the other hand, the Bayes risk of the UMVU estimator is

$$\mathrm{E}_p[\mathrm{E}_{X|p}[(X/n - p)^2]].$$

Performing the inner expectation yields

$$\mathrm{E}_{X|p}[(X/n - p)^2] = \frac{p(1-p)}{n}$$

and consequently

$$\mathrm{E}_p[\mathrm{E}_{X|p}[(X/n - p)^2]] = \frac{ab}{(a+b)(a+b+1)n}.$$

■

LC 4.2.2.b Suppose that the number of Bernoulli trials performed is a random variable N. The value $N = n$ was obtained when a prespecified number x of successes was observed so that we have the negative binomial sample model

$$P(N = n) = \binom{n-1}{k-1} p^x (1-p)^{n-x},\ n \geq x.$$

Calculate the Bayes risk of the Bayes estimator and the UMVU estimator of p.

Here, as far as anyone can tell, there's no nice solution. The Bayes estimator is still

$$\delta_\pi(N) = \frac{a+x}{a+b+N}$$

and now the UMVU estimator is $(x-1)/(N-1)$.

Again, we can find the Bayes risk of an estimator d by computing

$$\mathrm{E}_p[\mathrm{E}_{N|p}[(d-p)^2]].$$

However, now both our estimators d have N in the denominator and it's hard to take the expectation of $1/N$ or $1/N^2$. It's still possible to come up with *something* for $E_{N|p}[(d - p)^2]$, but it'll involve an infinite sum. ■

LC 4.2.2.c Calculate the mean squared errors of all three estimators under each model. If it is unknown which sampling mechanism generated the data, which estimator do you prefer overall?

For the first model, we already calculated the mean squared errors as $\mathrm{E}_{X|p}[(d-p)^2]$. For the third estimator $d = \frac{X-1}{n-1}$, the mean squared error is given by

$$\mathrm{E}_{X|p}[((X-1)/(n-1) - p)^2] = \frac{np(1-p) + (1-p)^2}{(n-1)^2}.$$

Under the second model, the mean squared errors will all be ugly infinite sums.

In any case, even without being able to compare mean squared errors, we might note that the form of the Bayes estimator is the same regardless of the data generating mechanism and prefer it for that reason alone. ■