**Solutions Homework 5**
March 16, 2018

**Solution to Exercise 5.5.1 (a)** It is well known that to minimize over $a$ the expression $E[(X-a)^2]$, one sets $a = \mu = E[X]$. It’s easy to show:

$$E[(X-a)^2] = E[(X-\mu + \mu-a)^2] = E[(X-\mu)^2] + 2(\mu-a)E[X-\mu] + (\mu-a)^2 = E[(X-\mu)^2] + (\mu-a)^2,$$

and setting $\mu = a$ gives the minimum value of $E[(X-\mu)^2]$. Just apply this result with $X = g(\theta)$ where the distribution of $\theta$ is given by the posterior.

**(b)** We wish to minimize the expression

$$\rho(x,d) = \int \left( \frac{d - g(\theta)}{u(\theta)} \right)^2 f(x|\theta)\pi(\theta)d\theta$$

As long as

$$\int f(x|\theta)\pi(\theta)u^{-2}(\theta)d\theta < \infty$$

then we could normalize $f(x|\theta)\pi(\theta)u^{-2}(\theta)$ to be a probability density, and pretend it is the posterior. Then, the answer would be the posterior mean with respect to this modified posterior. That is, the Bayes estimator would be

$$\delta^*(x) = \frac{\int g(\theta)f(x|\theta)\pi(\theta)u^{-2}(\theta)d\theta}{\int f(x|\theta)\pi(\theta)u^{-2}(\theta)d\theta}.$$ 

We would also need

$$\int g(\theta)^2 f(x|\theta)\pi(\theta)u^{-2}(\theta)d\theta < \infty$$

for this to make sense (else the Bayes risk is always infinite).

Note that we could also consider adjusting the prior by incorporating the factor $u^{-2}(\theta)$ in the prior. However, this may lead to an improper prior, which can create difficulties. The requirement in the last displayed inequality is simply that the “modified” posterior is proper. However, we will sometimes consider the modification of the prior when that makes sense (see part (d)(i) below).

**(c)** Just replace $u(\theta)$ by $g(\theta)$ in the above.

**(d) (i)** Let’s switch to the precision parameter $\lambda = 1/\sigma^2$ as done right before equation (3.169) (should be (5.169)). This means that we have an
extra factor of \( \lambda \) in the formula for the unnormalized posterior expected loss, i.e. we want to minimize

\[
\rho(x, d) = \int_0^\infty \int_{-\infty}^{\infty} (\mu - d)^2 \lambda f(x|\mu, \lambda) \pi(\mu|\lambda) \pi(\lambda) \, d\mu d\lambda.
\]

If we absorb that constant into \( \pi(\lambda) \), then it is like effective changing it from a \( \text{Gamma}(\alpha, \beta) \) prior to a \( \text{Gamma}(\alpha + 1, \beta) \) prior since the factor \( \lambda^{\alpha-1} \) in the \( \text{Gamma}(\alpha, \beta) \) prior will become \( \lambda^\alpha \). Thus, we can compute the Bayes estimator under squared error loss but use this modified prior. However, in the derivation of (3.176) it was noted that the posterior mean of \( \mu \) doesn’t depend on the prior parameters for \( \lambda \), so (3.176) also gives the Bayes estimate under this weighted squared error loss.

Note that we do not need any restrictions on the prior parameters or on \( \mu \) or \( \lambda \) in order for this to hold.

(ii) The marginal posterior for \( \mu \) is given by (3.178) where \( \eta \) is given in (3.173) and \( \hat{\mu} \) is given in (3.176). Both are posterior parameters determined from the data and prior parameters. As noted in the comment that follows this formula, this is the density for a shifted and rescaled \( t \) distribution with \( n + 2\alpha \) degrees of freedom, provided the latter is an integer, which doesn’t really matter for the formula for the density (and we can extend the \( t \) family to allow a continuous degrees of freedom parameter). The pdf for the \( t \) distribution with \( \nu \) degrees of freedom is given by

\[
f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\Gamma\left(\frac{\nu}{2}\right)}} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}.
\]

We only need \( \nu > 0 \) for this to be pdf. Thus, we can write down the p.d.f by inspection:

\[
\pi(\mu|x) = \frac{\sqrt{\eta(n+c)(n+2\alpha)\Gamma\left(\frac{n+2\alpha+1}{2}\right)}}{\sqrt{(n+2\alpha)\pi\Gamma\left(\frac{n+2\alpha}{2}\right)}} \left(1 + \frac{\eta(n+c)(n+2\alpha)(\mu - \hat{\mu})^2}{n + 2\alpha}\right)^{-(n+2\alpha+1)/2}
\]

\[
= \frac{\sqrt{\eta(n+c)\Gamma\left(\frac{n+2\alpha+1}{2}\right)}}{\sqrt{\pi\Gamma\left(\frac{n+2\alpha}{2}\right)}} \left(1 + \eta(n+c)(\mu - \hat{\mu})^2\right)^{-(n+2\alpha+1)/2}.
\]

(Note: To help avoid errors, let’s check that the last expression has the right units. since it is a Lebesgue pdf, it should have units of \( 1/\mu \). Note that the \((1 + ...)^{-(n+2\alpha+1)/2}\) factor is dimensionless since \( \eta \) has units of \( 1/\mu^2 \), and with
the $\eta^{1/2}$ factor in the first factor (the numerator of the normalizing constant fraction) we do indeed have units of $1/\mu$.

To finish up, we only need to compute the posterior mean of $\mu$ with respect to this posterior density. One can write out the integral, but doesn’t have any obvious simplification. I put it into Mathematica and it returned it to me as a scaled hypergeometric function, which is OK, but not particularly enlightening.

In order for this to be meaningful, we need for $\mu$ to have finite fourth moments under the posterior. In order that
$$\int_{-\infty}^{\infty} y^4 \frac{1}{(1 + ay^2)^b} dy < \infty,$$
(where $a > 0$), we need that $2b > 5$, so for this exercise to be meaningful, we need
$$n + 2\alpha > 4.$$

(iii) Our loss function is
$$L(\mu, \sigma^2, d) = \left( \frac{d - \sigma^2}{\sigma^2} \right)^2 = (d\lambda - 1)^2,$$
where again we have switched to the precision parameter in the last expression. The Bayes estimator will be found by minimizing the posterior expected loss
$$\bar{\rho}(\mu, d) = \int_{0}^{\infty} (d\lambda - 1)^2 \pi(\lambda|d) d\lambda.$$
Expanding the quadratic expression in the integrand and minimizing over $d$ gives the Bayes estimator
$$\delta^*(\mu) = \frac{\int_{0}^{\infty} \lambda \pi(\lambda|d) d\lambda}{\int_{0}^{\infty} \lambda^2 \pi(\lambda|d) d\lambda}.$$
We have from equation (3.172) that $\pi(\lambda|d)$ is a $Gamma(n/2 + \alpha, \eta)$ density. The first and second moments of $Gamma(\alpha, \beta)$ are $\alpha\beta$ and $\alpha(\alpha + 1)\beta^2$, respectively. Plugging in the appropriate parameter values we get
$$\delta^*(\mu) = \frac{2}{(n + 2\alpha + 2)\eta},$$
where again $\eta$ is given in (3.173).
We don’t need any restrictions on \( n \) or the prior parameters for this to be meaningful since, when reexpressed in terms of \( \lambda \), we are only dealing with second moments of \( \lambda \) in the loss function, and it has a prior and posterior Gamma distribution, which has moments of all orders.

**Solution to Exercise 5.6.7 3 points** Starting with squared error loss, we try to find a least favorable sequence of priors among the Beta priors, since this is a conjugate family for the Binomial \( p \). So assuming \( p \sim Beta(\alpha, \beta) \), we get the posterior

\[
\pi(p|x) \propto f(x|p)\pi(p) \\
\propto p^{\alpha+x-1}(1-p)^{\beta+n-x-1},
\]

which is clearly a \( Beta(\alpha + x, \beta + n - x) \) distribution. The Bayes estimate under squared error loss is the posterior mean, so we get the Bayes rule

\[
\delta_{\alpha\beta}(x) = \frac{x + \alpha}{n + \alpha + \beta}.
\]

Our least favorable sequence of priors will most likely be found by looking for an extended Bayes equalizer rule. To find this, we first compute the (frequentist) risk:

\[
R(p, \delta_{\alpha\beta}) = MSE(p, \delta_{\alpha\beta}) \\
= Bias^2(p, \delta_{\alpha\beta}) + Var_p(\delta_{\alpha\beta}) \\
= \left[ p - \frac{np + \alpha}{n + \alpha + \beta} \right]^2 + \frac{np(1-p)}{(n + \alpha + \beta)^2} \\
= \frac{(\beta p + \alpha(p - 1))^2 + np(1-p)}{(n + \alpha + \beta)^2} \\
= \frac{((\beta + \alpha)^2 - n)p^2 + (n - 2\alpha(\alpha + \beta))p + \alpha^2}{(n + \alpha + \beta)^2}.
\]

Looking at the above, we see that the denominator is independent of \( p \), and the numerator is a quadratic function of \( p \). If we can choose \( \alpha \) and \( \beta \) so that the coefficients of \( p \) and \( p^2 \) disappear, then the numerator will be independent of \( p \) and we will have a Bayes equalizer rule, which will be minimax. Thus, we get the system of equations

\[
(\beta + \alpha)^2 = n \\
2\alpha(\alpha + \beta) = n
\]
Rather than working through the algebra, let’s use some intuition. Since the risk will be constant, the prior should not “favor” \( p > 1/2 \) or \( p < 1/2 \), i.e., should be symmetric. Thus, we can guess that \( \alpha = \beta \). With this guess, the first equation gives
\[
\alpha = \beta = \sqrt{n}/2,
\]
and one can verify that then te second equation is solved. With this prior we get the Bayes rule
\[
\delta(X) = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}.
\]

The risk (MSE) is then
\[
MSE(p, \delta) = \frac{n}{4(n + \sqrt{n})^2}.
\]

Note that the MSE for the “usual” estimator \( \hat{p} = X/n \), which is the UMVUE, is \( MSE(\hat{p}, p) = p(1-p)/n \). This has a maximum value at \( p = 1/2 \) of \( 1/(4n) \), which is larger than the risk of the minimax estimator above. Note that the minimax estimator is biased, and in fact all of its MSE comes from its bias when \( p = 0 \) or \( p = 1 \).

Now we consider the weighted squared error loss function given in the exercise. Looking at the risk for the UMVUE we have
\[
\frac{MSE(p, \hat{p})}{p(1-p)} = \frac{p(1-p)/n}{p(1-p)} = 1/n,
\]
and so it is an equalizer rule for this loss function. We would conjecture that it is the minimax rule. In order to apply the theory, we need to show that it is Bayes or extendend Bayes. Again using a \( Beta(\alpha, \beta) \) prior, we have that the Bayes estimator will minimize over \( d \) the expression
\[
\int_0^1 (d-p)^2 p^{\alpha+\gamma-2}(1-p)^{\beta+n-\gamma-2}dp.
\]

This is the same expression we would minimize if we used squared error loss with a \( Beta(\alpha - 1, \beta - 1) \) prior, and we see that the Bayes estimator is
\[
\delta_{\alpha\beta}(x) = \frac{x + \alpha - 1}{n + \alpha + \beta - 2}.
\]
(Note in general that when using weighted squared error loss, one can generally absorb the weight into the prior and the Bayes rule will be the posterior
mean under the modified prior.) Clearly the choice $\alpha = \beta = 1$ results in the Bayes estimator being $\hat{p}$, so we see that the UMVUE is the minimax rule under this loss. Note that we don’t have to compute the risk, since we could produce an equalizer rule, but one could do the analogous calculations to those done in the first part of the problem and arrive at the same conclusion.