STAT 830

Problems: Assignment 5

1. For a sample of size n from the Uniform $[0,\theta]$ distribution, compute the posterior law of θ if you use the prior $\pi(\theta) = 1/\theta$. What is the posterior Bayes risk? What is the Bayes estimator for squared loss.

   Likelihood times prior is

$$\prod_{i=1}^{n} \frac{1(0 < X_i < \theta)}{\theta} \times \frac{1}{\theta}$$

   so the posterior is 0 on $\theta \le \max\{X_i\} = X_{(n)}$. For $\theta > X_{(n)}$ the posterior is proportional to $1/\theta^{n+1}$. Divide by the normalizing constant

$$\int_{X_{(n)}}^{\infty} dt/t^{n+1}$$

   to get

$$\pi(\theta|X_1,\ldots,X_n) = \frac{nX_{(n)}^n}{\theta^{n+1}} 1(\theta > X_{(n)}).$$

   Here $X_{(n)} = \max\{X_1,\ldots,X_n\}$ is the largest order statistic. This is a Pareto distribution. The posterior Bayes risk of an estimate $\hat\theta$ is

$$nX_{(n)}^n \int_{X_{(n)}}^{\infty} (\hat\theta - \theta)^2/\theta^{n+1}\, d\theta$$

   which is

$$\frac{(n^2-3n+2)\hat\theta^2 - 2(n^2-2n)X_{(n)}\hat\theta + (n^2-n)X_{(n)}}{?}$$

   This is minimized by

$$\hat\theta = \frac{n}{n-1} X_{(n)}.$$

The Bayes risk of $\hat{\theta}$ would be

$$\int_0^\infty \mathrm{E}_\theta \left\{ (\hat{\theta} - \theta)^2 \right\} \frac{1}{\theta} \, d\theta.$$

The density of $X_{(n)}$, given θ,is

$$h(t) = n \frac{t^{n-1}}{\theta^n} 1(0 < t < \theta)$$

so the risk of our Bayes estimator is

$$\int_0^\theta (t - \theta)^2 n \frac{t^{n-1}}{\theta^n} \, dt = \frac{\theta^2 (n^2 - n + 2}{(n+2)(n+1)(n-1)^2}$$

(I used `Maple`) so the Bayes risk becomes

$$\int_0^\infty \frac{\theta^2 (n^2 - n + 2}{(n+2)(n+1)(n-1)^2 t\theta} \, d\theta = \infty.$$

This problem is one of those where every procedure is technically Bayes because the prior Bayes risk is necessarily infinite for this prior. But the posterior Bayes risk is finite for $n > 1$ and you can minimize the posterior risk.

2. In the same problem use the factorization theorem to find a minimal sufficient statistic. Show that this statistic is complete.

The likelihood is

$$L(\theta) \prod_{i=1}^n \frac{1(0 < X_i < \theta)}{\theta}$$

which is

$$L(\theta) = \prod_{i=1}^n 1(0 < X_i) \times \frac{1(\max_i \{X_i\} < \theta)}{\theta^n}.$$

The first term is $h(X)$ and the second depends only on the statistic $X_{(n)}$. Notice that

$$\frac{L(\theta)}{L(1)} = \frac{1(\max_i \{X_i\} < \theta)}{\theta^n 1(max_i \{X_i\} < 1)}$$

which is only a function of $\max_i\{X_i\}$. This shows that the statistic is minimal sufficient.

To see completeness: the density of $X_{(n)}$ is

$$f(t) = \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1} 1(0 < t < \theta).$$

If $\mathrm{E}(h(X_{(n)}) = 0$ for all θ then

$$n\int_0^\theta h(t)t^n\,dt/\theta^n \equiv 0$$

which meaans that

$$\int_0^\theta h(t)t^n\,dt \equiv 0.$$

The right hand side is differentiable so the left hand side is too. The derivative is 0 and so for almost every t we must have $h(t) = 0$. If we change h on a set of Lebesgue measure 0 we don't change the integral so this is the best possible conclusion.

3. Suppose I observe a sample of size n from the Poisson distribution with mean λ. Give λ a Gamma(α, β) distribution. Compute the Bayes estimator for squared error loss, find the Bayes risk, and either find a minimax estimate or explain why you can't.

Some people took β to be a rate parameter rather than a scale parameter – replacing my β below by $1/\beta$; that choice makes for tidier formulas but doesn't match the parameterization I used in other homework problems. I was happy with either choice.

The prior times likelihood is

$$\frac{1}{\Gamma(\alpha)}\frac{\lambda^{\alpha-1}}{\beta^\alpha}\exp(-\lambda/\beta)\lambda^{\sum_i X_i}exp(-n\lambda)/\prod X_i!$$

which is proportional to

$$\lambda^{\alpha+T-1}\exp(-\lambda(n+1/\beta))$$

where $T = \sum X_i$. This means the posterior is also Gamma but now with parameters $\alpha + T$ and $n + 1/\beta$. The posterior mean is

$$\hat{\lambda} = \frac{\alpha + T}{n + 1/\beta}$$

The risk of this estimator is its variance plus the square of its bias:

$$\frac{\text{Var}(T)}{(n + 1/\beta)^2} + \left(\frac{\alpha + n\lambda}{n + 1/\beta} - \lambda\right)^2$$

which simplifies to

$$\frac{\lambda^2 + \beta(n\beta - 2\alpha)\lambda + \alpha^2\beta^2}{(n\beta +)^2}$$

The Bayes risk is obtained by multiplying by the Gamma(α, β) density and integrating from $\lambda = 0$ to ∞. Maple gives me

$$\frac{\alpha\beta^2}{1 + n\beta}.$$

The risk function of a minimax Bayes estimator is constant but our risk function is quadratic and it is not possible to make the coefficient of the quadratic term be 0 so this method cannot find a minimax estimate. I think it is a theorem that there is no useful minimax estimator because every estimator has risk going to $infty$ as $\lambda \to \infty$.

4. This problem concerns the loss function for which a posterior mode would be a Bayes estimate. The parameter space Θ is finite and

$$L(\theta, d) = 1(\theta \neq d).$$

Show that the posterior mode is the Bayes estimate. Assume that the data X has a discrete distribution for each $\theta \in \Theta$.

An estimator would be a function $\delta(X)$ whose values are in Θ. To compute the risk function of $\delta(X)$ we would compute

$$\begin{aligned} R(\theta_i, \delta) &= \text{E}_{\theta_i}(L(\theta_i, \delta(X))) \\ &= P_{\theta_i}(\delta(X) \neq \theta_i) \end{aligned}$$

If we had a prior π we could think of π has being a vector $\pi_1, \ldots, \pi_k$ with $\pi = P(\theta = \theta_i)$. Then the Bayes risk would be

$$r(\pi, \delta) = \sum_{i=1}^{k} \pi_i P_{\theta_i}(\delta(X) \neq \theta_i).$$

The way we find Bayes procedures is to recognize that there are two expected values here – one is the sum over i and the other is the P_{θ_i}. For a Bayesian it is better to write

$$r(\pi, \delta) = \sum_{i=1}^{k} \pi_i P(\delta(X) \neq \theta_i | \theta = \theta_i).$$

Then we apply Bayes theorem to the joint distribution of X and θ and get

$$r(\pi, \delta) = \mathrm{E}\left[\mathrm{E}(1(\delta(X) \neq \theta)|X)\right]$$

The quantity in the inside

$$\mathrm{E}(1(\delta(X) \neq \theta)|X)$$

is the *posterior Bayes risk* and we choose $\delta(X)$ to minimize this risk separately for each X. The conditional expectation in question treats θ as random with the conditional probability that $\theta = \theta_i$ being the posterior probability which we would write as $\pi(\theta_i|X)$. So the posterior Bayes risk is

$$\mathrm{E}(1(\delta(X) \neq \theta)|X) = 1 - \mathrm{E}(1(\delta(X) = \theta)|X)$$

and this last can be written as

$$1 - \sum_{i=1}^{k} \pi(\theta_i|X) 1(\delta(X) = \theta_i)$$

If I take $\delta(X) = \theta_j$ for some particular j then the Bayes risk works out to be $1 - \pi(\theta_j|X)$. To make this small, make the thing being subtracted large – that is, choose j to maximize the posterior probability of θ_j. That is the posterior mode.

5. I discussed Bayesian testing problems a bit. In this problem you are testing $\mu = 0$ against $\mu \neq 0$ based on the sample mean $\bar{X} \sim N(\mu, 1/n)$. Your prior is $P(\mu = 0) = 1/2$ and given $\mu \neq 0$ the prior has a $N(0, \tau^2))$ distribution. Plot the P-value for the usual test which rejects for large values of $|\sqrt{n}\bar{X}|$ as a function of the observed value x of $\bar{X}$. On the same plot graph $P(\mu = 0|\bar{X} = x)$ for a variety of values of τ. Try some τs which are quite small, and some which are quite large to get an idea of the effect of the prior variance on the conclusion.

The standard hypothesis test of $\mu = 0$ against $\mu \neq 0$ rejects for large values of $T = |\sqrt{n}\bar{X}|$ so

$$P = Prob(N(0,1) > t)|_{t=T} = 2(1 - \Phi(T))$$

The posterior probability that $\mu = 0$ given the data is

$$P(\mu = 0|\bar{X}) = \frac{\frac{1}{2}f(\bar{X}|\mu = 0)}{\frac{1}{2}f(\bar{X}|\mu = 0) + \frac{1}{2}\int(f(\bar{X}|\mu)\pi(\mu)\,d\mu}$$

In this formula $pi(\mu)$ is the conditional prior density of μ given $\mu \neq 0$ which is the $N(0, \tau^2)$ density. Given mu the random variable $\bar{X}$ is $N(\mu, 1/n)$ so $\bar{X} - \mu$ is $\mathrm{N}(0, 1/n)$. Because this conditional distribution does not depend on μ we see that μ and $\bar{X} - \mu$ are independent. The integral above is the density of

$$\bar{X} = (\bar{X} - \mu) + \mu$$

which has a $\mathrm{N}\big(0, \frac{1}{n} + \tau^2\big)$ distribution. Cancelling some common factors we find

$$P(\mu = 0|\bar{X}) = \frac{1}{1 + \frac{1}{\sqrt{\tau^2 + 1/n}} \exp\left(\frac{n\bar{X}^2}{2} - \frac{\bar{X}^2}{\tau^2 + 1/n}\right)}$$

which can be rewritten in many ways.

You have to pick an n to plot this for and I suggested $n = 1$; you need to plot $2(1 - \Phi(x))$ and

$$\frac{1}{1 + \frac{1}{\sqrt{1+\tau^2}} \exp\left(-\frac{\tau^2 x^2}{1+\tau^2}\right)}$$

against x but I won't produce plots here.

6. Suppose $X_1, \ldots, X_m$ are iid $N(\mu, \sigma^2)$ and $Y_1, \ldots, Y_n$ are iid $N(\chi, \tau^2)$. Assume the Xs are independent of the Ys.

   (a) Find complete and sufficient statistics.

   The log likelihood is

$$-(n+m)\log(2\pi)/2 - m\log(\sigma) - n\log(\tau) - \frac{\sum X_i^2}{2\sigma^2} - \frac{\sum Y_i^2}{2\tau^2} + \frac{m\bar{X}\mu}{\sigma^2} + \frac{n\bar{Y}\chi}{\tau^2} - \frac{m\mu^2}{2\sigma^2} - \frac{n\chi^2}{2\tau^2}$$

   Since this is the form of an exponential family it follows that

$$S = (\bar{X}, \sum X_i^2, \bar{Y}, \sum Y_i^2)$$

   is complete and sufficient. There are many acceptable alternative one to one functions of this vector. A correct answer will check that the range of the coefficients of the sufficient statistics fills out a 4 dimensional hyper-rectangle. As I have written it the coefficients of the 4 statistics are

$$\big(m\mu/\sigma^2, -1/(2\sigma^2), n\chi/\tau^2, -1/(2\tau^2)\big)$$

   As μ and χ each vary over $\mathbb{R}$ and σ and τ vary over the positive reals this coefficient vector varies over the 4 dimensional hyper-rectangle

$$\mathbb{R} \times (-\infty, 0) \times \mathbb{R} \times (-\infty, 0).$$

   This proves that our vector S is complete and sufficient.

   (b) Find UMVUE's of $\mu - \chi$ and σ^2/τ^2.

   Since $E(\bar{X} - \bar{Y}) = \mu - \chi$ the Lehmann-Shceffé theorem proves that the UMVUE of $\mu - \chi$ is $\bar{X} - \bar{Y}$.
   The random variable $[s_x^2/\sigma^2]/[s_y^2/\tau^2]$ has an F distribution so that

$$E(s_x^2/s_y^2) = c\sigma^2/\tau^2$$

   If we find c then the estimate $c^{-1}s_x^2/s_y^2$ will be the UMVUE by Lehmann-Scheffé.

But

$$
\begin{aligned}
E([s_x^2/\sigma^2]/[s_y^2/tau^2]) &= E(\chi^2_{m-1}/(m-1))E((n-1)/\chi^2_{n-1}) \\
&= \int_0^\infty \frac{n-1}{u}\frac{1}{\Gamma((n-1)/2)}\left(\frac{u}{2}\right)^{(n-1)/2-1} e^{-u/2}d(u/2) \\
&= \int_0^\infty \frac{n-1}{2y}\frac{1}{\Gamma((n-1)/2)} y^{(n-1)/2-1}e^{-y}dy \\
&= \frac{n-1}{2\Gamma((n-1)/2)}\int_0^\infty y^{(n-3)/2-1}e^{-y}dy \\
&= \frac{(n-1)\Gamma((n-3)/2)}{2\Gamma((n-1)/2)} \\
&= \frac{n-1}{n-3}
\end{aligned}
$$

So the UMVUE is

$$\frac{n-3}{n-1}\frac{s_x^2}{s_y^2}$$

(c) Now suppose you know that $\sigma = \tau$. Find UMVUE's of $\chi - \mu$ and of $(\chi - \mu)/\sigma$. (You have already found the UMVUE for σ^2.)

You should check as I did in a) that for the smaller model the vector of complete sufficient statistics is the 3 vector

$$(\bar{X}, \bar{Y}, \sum(X_i - \bar{X})^2 + \sum(Y_i - \bar{Y})^2)$$

This is one to one with the vector

$$(\bar{X}, \bar{Y}, \sum_i X_i^2 + \sum_j Y_j^2)$$

which is therefore also complete and sufficient. According to Lehmann-Scheffé $\bar{Y} - \bar{X}$ is still the UMVUE of $\chi - \mu$. Since $(\bar{X}, \bar{Y})$ is independent of

$$s^2 = [\sum(X_i - \bar{X})^2 + \sum(Y_i - \bar{Y})^2]/(n+m-2)$$

we find

$$E[(\bar{Y} - \bar{X})/s] = \frac{\chi - \mu}{\sigma}E(\sigma/s)$$

Now $(n+m-2)s^2/\sigma^2$ has a χ^2_{n+m-2} distribution so

$$\begin{aligned} E(\sigma/s) &= \sqrt{n+m-2}\int_0^\infty u^{-1/2}\frac{1}{\Gamma((n+m-2)/2)}\left(\frac{u}{2}\right)^{(n+m-2)/2-1} e^{-u/2}d(u/2) \\ &= \sqrt{n+m-2}\int_0^\infty \frac{1}{2\Gamma((n+m-2)/2)}y^{(n+m-2)/2-1-1/2}e^{-y}dy \\ &= \frac{(n+m-2)\Gamma((n+m-3)/2)}{2\Gamma((n+m-2)/2)} \end{aligned}$$

It follows that the UMVUE of $(\chi-\mu)/\sigma$ is

$$\frac{2\Gamma((n+m-2)/2)}{(n+m-2)\Gamma((n+m-3)/2)}\frac{\bar{Y}-\bar{X}}{s}$$

(d) Now suppose σ and τ are unknown but that you know that $\mu=\chi$. Prove there is no UMVUE for μ. (Hint: Find the UMVUE if you knew $\sigma/\tau = a$ with a known. Use the fact that the solution depends on a to finish the proof.)

(e) Why doesn't the Lehmann-Scheffé theorem apply?

We are analyzing the model $X_1,\ldots,X_m$ iid $N(\mu,\sigma^2)$ and independently $Y_1,\ldots,Y_n$ iid $N(\mu,\tau^2)$. I suggested studying the submodel $X_1,\ldots,X_m$ iid $N(\mu,\sigma^2)$ and independently $Y_1,\ldots,Y_n$ iid $N(\mu,a^2\sigma^2)$. For this model the complete sufficient statistics are

$$\sum X_i^2 + \sum Y_i^2/a^2$$

and

$$\sum X_i + \sum Y_i/a^2$$

The UMVUE of μ is then

$$\hat\mu_a \equiv (\sum X_i + \sum Y_i/a^2)/(m+n/a^2)$$

Now if there were a UMVUE $\hat\mu$ for the big model we would have to have

$$\text{Var}(\hat\mu) \le \text{Var}(\hat\mu_a)$$

for all σ^2, τ^2 and μ. (This uses the fact that $\hat\mu_a$ is unbiased in the full model for any choice of a.) In particular, this would have to be true for each τ and σ for which $\tau = a\sigma$. But then $\hat{\mu}$ would be UMVUE in the small model. Since UMVUEs are unique in the small model (Lehmann-Scheffé and Rao-Blackwell) we find that

$$\hat{\mu} = \hat{\mu}_a$$

Since this can't be true for two different values of a there must not exist a UMVUE in the big model.
I got a lot of vague proofs which said it was obvious from the fact that μ_a depended on a but I wanted more detail. To apply Lehmann Scheffé in this model you would need a vector of complete sufficient statistics. For this 3 parameter model the log-likelihood is

$$-\frac{\sum X_i^2}{2\sigma^2}-\frac{\sum Y_i^2}{2\tau^2}+\frac{\mu}{\sigma^2}\sum X_i+\frac{\mu}{\tau^2}\sum Y_i+c(\mu,\sigma,\tau)-(n+m)\log(2\pi)$$

This is the exponential family form but it involves 3 parameters and 4 statstics. If you fix the value of 3 of the coefficients you can compute the 4th coefficient so the set of possible values does not contain a 4 dimensional hyper-rectangle.
Many people did this whole problem without ever checking the conditions of the Lehmann-Scheffé theorem.

7. Suppose $X_1, \ldots, X_n$ iid Poisson(λ). Find the UMVUE for λ and for $1 - \exp(-\lambda) = P(X_1 \neq 0)$.

   $\bar{X}$ is complete and sufficient and unbiased for λ. Hence the UMVUE of λ is $\bar{X}$. IF

   $$T = 1(X_1 \neq 0)$$

   we see that T is unbiased for $\phi = 1 - \exp(-\lambda)$ then the UMVUE of ϕ is

   $$E(T|\bar{X})$$

To compute this we note

$$E(T|\sum X_i = x) = P(X_1 \neq 0|\sum X_i = x)$$
$$= 1 - P(X_1 = 0|\sum X_i = x)$$
$$= 1 - \frac{P(X_1 = 0, X_1 + \cdots + X - n = x)}{P(\sum X_i = x)}$$
$$= 1 - \frac{P(X_1 = 0)P(X_2 + \cdots + X_n = x)}{P(\sum X_i = x)}$$

Remember that $X_1 + \cdots + X_n$ is Poisson$(n\lambda)$ and that $X_2 + \cdots + X_n$ is Poisson$((n-1)\lambda)$ to get

$$E(T|\sum X_i = x) = 1 - \frac{e^{-\lambda}\exp(-(n-1)\lambda)((n-1)\lambda)^x/x!}{exp(-n\lambda)(n\lambda)^x/x!}$$
$$= 1 - \left(\frac{n-1}{n}\right)^x$$

Conditioning on $\sum X_i = x$ is the same as conditioning on $\bar{X} = x/n$ we see that the UMVUE of ϕ is

$$1 - (1 - 1/n)^{n\bar{X}}$$

8. Suppose $\phi(X)$ is a test function and $S(X)$ is a sufficient statistic for some model. Show that
$$E(\phi(X)|S)$$
is a test function and compare its power and level to that of $\phi(X)$.

Since $0 \leq \phi(X) \leq 1$ we find

$$0 = E(0|S) \leq E(\phi(X)|S) \leq E(1|S) = 1$$

so that $E(\phi(X)|S)$ is between 0 and 1. Since S is sufficient this function is a statistic and so a test function. Its power function is

$$\pi(\theta) = E_\theta(E(\phi(X)|S)) = E_\theta(\phi(X))$$

which is the same as the power function of ϕ.

Many papers never really used the fact that S is assumed to be sufficient. If S is not sufficient then usually $\mathrm{E}(\phi(X)|S)$ would depend on θ. A test function must not – it must be a statistic.

9. Suppose $X_1, \ldots, X_n$ are iid exponential(λ).

   (a) Find the exact confidence levels of 95% intervals based on normal approximations to the distributions of the pivots $T_1 = \bar{X}/\lambda$, $T_2 = 1/T_1$, and $T_3 = \log(T_1)$ for n=10, 20 and 40.

   The standard normal approximation to T_1 would be

$$\sqrt{n}(T_1 - 1) \sim N(0,1)$$

   For T_2 we write $T_2 = f(T_1)$ with $f(x) = 1/x$ and get

$$\sqrt{n}(T_2 - 1) \sim \sqrt{n} f'(1)(T_1 - 1) \sim N(0,1)$$

   Finally for T_3 we use $f(x) = \log(x)$ and again get

$$\sqrt{n} T_3 \sim N(0,1)$$

   If $z_0.025$ is the upper 2.5% point for a normal curve then our intervals are based on

$$\{\lambda : |T_1 - 1| \le z_{0.025}/\sqrt{n}\}$$

$$\{\lambda : |T_2 - 1| \le z_{0.025}/\sqrt{n}\}$$

$$\{\lambda : |T_3| \le z_{0.025}/\sqrt{n}\}$$

   We invert these inequalities to get corresponding intervals

$$\frac{\bar{X}}{1 + z_{0.025}/\sqrt{n}}, \frac{\bar{X}}{1 - z_{0.025}/\sqrt{n}}$$

$$\bar{X}(1 + z_{0.025}/\sqrt{n}), \bar{X}(1 - z_{0.025}/\sqrt{n})$$

   and

$$\bar{X}\exp[-z_{0.025}/\sqrt{n}], \bar{X}\exp[z_{0.025}/\sqrt{n}]$$

To evaluate the coverage probabilities we remember that

$$\bar{X}/\lambda$$

has a Gamma$(n, 1/n)$ distribution. I get the following table

n	T_1	T_2	T_3
10	0.9549216	0.9035129	0.9410225
20	0.9528124	0.9255531	0.9454931
40	0.9514818	0.9375773	0.9477421

The pivot T_1 works best.

Some students worked harder and computed the exact mean and variance of the pivots T_2 and T_3 which lead to much more accurate normal approximations. This was good.

(b) Find the shortest exact 95% confidence interval based on T_1; get numerical values for n=10, 20 and 40.

The interval in question

$$\bar{X}/\gamma_1 \le \lambda \le \bar{X}/\gamma_2$$

where γ_1 is an upper critical point for the Gamma$(n, 1/n)$ distribution and γ_2 a lower tail point and where the total area in the two tails is 0.05. The interval length is

$$\frac{1}{\gamma_2} - \frac{1}{\gamma_1}$$

I used R to evaluate the length of this interval on a grid and got the upper tail area to be 0.0373 and the lower tail area to be 0.0127 for $n = 10$, leading to $\gamma_1 = 1.630$ and $\gamma_2 = 0.429$. For $n = 20$ I get the corresponding areas to be 0.0337 and 0.0163 with $\gamma_1 = 1.446$ and $\gamma_2 = 0.583$. For $n = 40$ I get 0.0312 and 0.0188 with $\gamma_1 = 1.314$ and $\gamma_2 = 0.699$.

(c) Find the exact confidence level of 95% confidence intervals based on the chi-squared approximation to the distribution of deviance drop. Compare the results with the previous question based on length and coverage probabilities. Figure out how to make a convincing comparison. Which method is better?

The interval is

$$\left\{\lambda : 2(\ell(\hat{\lambda}) - \ell(\lambda)) \le \chi^2_{1,\alpha} = 3.84\right\}.$$

The deviance drop becomes

$$n\bar{X}/\lambda - n\log\bar{X}/\lambda - n = nT_1 - n\log(T_1) - n$$

and this is below 3.84 on an interval of T_1 values depending on n. This permits you to see that the χ^2 approximation is quite accurate (but a tiny bit liberal – the coverages are very slightly below 0.95).