Stat 300A Theory of Statistics

Problem session 2

Suyash gupta *Due on*

1: Some Examples

Table 1: Some Bayes Estimators

	Loss($L(\theta, d)$)	**Bayes Estimator**$\delta_\pi(x)$
1	$(d - g(\theta))^2$	$\mathbb{E}[g(\theta\vert x)]$
2	$w(\theta)(d - g(\theta))^2$	$\frac{\mathbb{E}[w(\theta)g(\theta)\vert x]}{\mathbb{E}[w(\theta)\vert x]}$
3	$\vert d - g(\theta)\vert$	$med\ \mathbb{L}[g(\theta)\vert x]$
4	$I\{\vert d - g(\theta)\vert > c\}$	$argmax_d P(d - c \leq \theta \leq d + c)$

2: Bayes Estimator for weighted squared error loss function

Suppose that $X \sim Poisson(\lambda)$ and consider the information normalized loss function $L_1(d, \lambda) = \frac{(d-\lambda)^2}{\lambda}$

(*a*) Find an expression for the Bayes estimator $\hat{\lambda_\pi}(x)$ for a prior π and loss function L_1
soln. Want to minimize

$$\mathbb{E}(w(\theta)(d - g(\theta))^2|x) \propto \int w(\theta)(d - g(\theta))^2\pi(\theta)p_\theta(x)dx$$

Absorb $w(\theta)$ into the posterior density and consider the new posterior density $\propto w(\theta)\pi(\theta)p_\theta(x)$. Now, the minimization problem is minimizing the risk for a squared error loss function w.r.t. the new posterior density which will be the corresponding new posterior mean. Hence, appropriately scaling the new posterior density to integrate to 1, we have the bayes estimate

$$\frac{\int g(\theta)w(\theta)\pi(\theta)p_\theta(x)dx}{\int w(\theta)\pi(\theta)p_\theta(x)dx} = \frac{\mathbb{E}[w(\theta)g(\theta)|x]}{\mathbb{E}[w(\theta)|x]}$$

(*b*) Evaluate this estimator, $\hat{\lambda_{a,b}}(x)$ say, for the (Gamma) conjugate priors $\pi_{a,b}$ where a is the shape parameter and b the scale parameter. Evaluate the risk function of $\hat{\lambda}_{a,b}$ for L_1.
soln. Simple calculations show that the Posterior distribution,

$$\lambda|X \sim \Gamma(x + a, \frac{b}{b+1})$$

So $\frac{1}{\lambda} \sim$ Inverse Gamma$(x + a, \frac{b}{b+1})$. From the expression in table above for the Bayes estimate for a squared error loss function, we have $\frac{b(x+a-1)}{b+1}$ as the required answer.

(c) Assume that $\hat{\lambda}_U(x) = x$ is admissible for squared error loss $L_0(d, \lambda) = (d-\lambda)^2$. Show that $\hat{\lambda}_U(x) = x$ is also admissible for L_1

soln. Use contradiction, if X is not admissible w.r.t. L_1, then clearly there exists an estimator $f(X)$ s.t.

$$\mathbb{E}(f(X)-\lambda)^2/\lambda \leq \mathbb{E}(X-\lambda)^2/\lambda$$

with strict inequality for some λ. Hence,

$$\mathbb{E}(f(X)-\lambda)^2 \leq \mathbb{E}(X-\lambda)^2$$

with strict inequality for some λ. Hence, contradiction.

(d) Hence show that $\hat{\lambda}_U$ is minimax for L_1. **soln.** Again use contradiction, if it is not minimax, then there exists an estimator $f(x)$ whose maximum risk is less than the maximum risk of X.

Note the risk function of X is constant 1 w.r.t. L_1. So, the risk function of $f(x)$ is strictly lower than 1 for all λ, hence contradicts the admissibility of X.

3: Uniform distribution and Gamma prior

$X_i|\theta \sim U(0, \theta)\ \ i = 1, 2, ..., n,$
$\frac{1}{\theta}|(a, b) \sim Gamma(a, b)$ a,b known

(a) Verify that the Bayes estimator for squared error loss will only depend on the data through $Y = max_i X_i$.
soln. $\theta \sim$ Inverse Gamma (a, b).

Posterior density is

$$\theta|\boldsymbol{X} \sim \frac{1}{b^a}\frac{1}{\theta^{(n+a+1)}}e^{-\frac{1}{\theta b}}\mathbf{1}\{max_i X_i < \theta\}$$

As posterior density depends on $\boldsymbol{X}$ through $Y = maxX_i$, so the posterior mean will also depend on $\boldsymbol{X}$ through $Y = maxX_i$.

(b) Show that $\mathbb{E}(\theta|y, a, b)$ can be expressed as $\mathbb{E}(\theta|y, a, b) = \frac{1}{b(n+a-1)}\frac{P(\chi^2_{2(n+a-1)} < 2/(by))}{P(\chi^2_{2(n+a)} < 2/(by))}$
soln. From a, it is sufficient to work only with the density of $Y = max_i X_i$, which is given by

$$g(y|\theta) = ny^{n-1}/\theta^n, \ \ 0 < y < \theta$$

Then calculate the posterior density of $\theta|y$, hence , the Bayes estimate is given by the correspondingh posterior mean

$$\mathbb{E}(\theta|y, a, b) = \frac{(\int_y^\infty \theta \frac{1}{\theta^{n+a+1}} e^{-1/(\theta b)} d\theta)}{(\int_y^\infty \frac{1}{\theta^{n+a+1}} e^{-1/(\theta b)} d\theta)}$$

Doing a simple change of variable $u = \frac{1}{\theta}$ yields the result.

4: Inadmissibility and minimaxity in Truncated Normal

Suppose X is distributed as $N(\theta, 1)$ where $\theta \geq 0$.

(a) Show that X is inadmissible.
soln. Let $X_+ = \max(X, 0)$. For any $\theta \geq 0$,

$$\mathbb{E}[(X_+ - \theta)^2] = \mathbb{E}[(X-\theta)^2 \mathbb{1}_{X \geq 0}] + \mathbb{E}[\theta^2 \mathbb{1}_{X<0}] < \mathbb{E}[(X-\theta)^2 \mathbb{1}_{X \geq 0}] + \mathbb{E}[(X-\theta)^2 \mathbb{1}_{X<0}] = \mathbb{E}[(X-\theta)^2],$$

so X_+ dominates X, and X is inadmissible.

(b) Show that X is a minimax estimate of θ for quadratic loss.
soln. The risk of X is $R(\theta, X) = \mathbb{E}_\theta[(X-\theta)^2] \equiv 1$. Suppose that $\delta(X)$ has maximum risk $1-\varepsilon$ for some $\varepsilon > 0$ and estimator $\delta(X)$; let $b(\theta) = \mathbb{E}_\theta[\delta(X)] - \theta$ be the bias of $\delta(X)$. Then, by the Cramer-Rao inequality (TPE Theorem 2.5.10), for any $\theta > 0$, $\mathrm{Var}_\theta(\delta(X)) \geq \frac{(1+b'(\theta))^2}{I(\theta)} = (1+b'(\theta))^2$ (since $I(\theta) = 1$ for the $\mathcal{N}(\theta, 1)$ model). So $1-\varepsilon \geq R(\theta, \delta(X)) \geq (1+b'(\theta))^2 + b(\theta)^2$ for all $\theta > 0$. This implies $1-\varepsilon \geq (1+b'(\theta))^2 \geq 1 + 2b'(\theta)$, so $b'(\theta) \leq -\frac{\varepsilon}{2}$ for all $\theta > 0$, and hence $\lim_{\theta\to\infty} b(\theta) = -\infty$. But this contradicts $1-\varepsilon \geq b(\theta)^2$ for all $\theta > 0$. Hence no such estimator $\delta(X)$ exists, the minimax risk is equal to 1, and X is minimax.

5: Bayes estimator for indicator loss function

Suppose that $X_1, ..., X_n$ are drawn independently from a common density $f(x-\theta)$, where $\theta \in R$ is an unknown location parameter and f belongs to the class F of densities with respect to lebesque measure on R. Assume that f is the density of a $N(0, 1)$ distribution. We are interested in finding the estimates of θ that are unlikely to be far from the truth, so for fixed $\epsilon > 0$, we adopt the loss function $L(d, \theta) = 1(|d-\theta| > \epsilon)$, defined for all $d, \theta \in R$.

(a) Find a Bayes estimator of θ under L and the prior $\theta \sim N(0, \tau^2)$ for known τ^2

Soln. The Bayes estimator is $\delta(X)$ for

$$\delta(x) = \mathrm{argmin}_\delta \mathbb{E}\left[\mathbb{1}\{|\delta - \Theta| > \varepsilon\} \mid X = x\right] = \mathrm{argmin}_\delta \mathrm{P}[|\delta - \Theta| > \varepsilon \mid X = x]$$

(TPE Theorem 4.1.1). The posterior density of Θ is given by

$$\begin{aligned} p(\theta|x) &\propto p(x, \theta) \\ &\propto \exp\left(-\frac{1}{2}\sum_i (x_i - \theta)^2\right)\exp\left(-\frac{\theta^2}{2\tau^2}\right) \\ &\propto \exp\left(-\frac{\theta^2}{2}\left(n + \frac{1}{\tau^2}\right) + \sum_i x_i\theta\right) \\ &\propto \exp\left(-\frac{n+\frac{1}{\tau^2}}{2}\left(\theta - \frac{\sum_i x_i}{n+\frac{1}{\tau^2}}\right)^2\right), \end{aligned}$$

hence $(\Theta|X = x) \sim \mathcal{N}\left(\frac{\sum_i x_i}{n+\frac{1}{\tau^2}}, \frac{1}{n+\frac{1}{\tau^2}}\right)$. The interval of length 2ε with maximum probability under a normal density is centered around the normal mean, hence $\delta(X) = \frac{\sum_i X_i}{n+\frac{1}{\tau^2}}$ is the Bayes estimator.

(b) Calculate the limit of the Bayes risk as $\tau \to \infty$.
soln. The risk function of $\bar{X}$ is

$$R(\theta) = \mathrm{P}_\theta[|\bar{X} - \theta| > \varepsilon] = \mathrm{P}[|Z| > \sqrt{n}\varepsilon] = 2\Phi\left(-\sqrt{n}\varepsilon\right),$$

where $Z \sim \mathcal{N}(0, 1)$. $R(\theta)$ is independent of θ since $\bar{X}$ is equivariant, hence the maximum risk of $\bar{X}$ is $2\Phi\left(-\sqrt{n}\varepsilon\right)$. The Bayes risk of the Bayes estimator $\delta(X)$ from part (b) is

$$r(\tau) = \mathrm{P}\left[|\delta(X) - \Theta| > \varepsilon\right] = \mathbb{E}\left[\mathrm{P}\left[|\delta(X) - \Theta| > \varepsilon \mid \Theta\right]\right].$$

Letting $Z \sim \mathcal{N}(0, 1)$,

$$\mathrm{P}\left[|\delta(X) - \theta| > \varepsilon \mid \Theta = \theta\right] = \mathrm{P}\left[\left|\frac{n}{n+\frac{1}{\tau^2}}\frac{\sum_i(X_i - \theta)}{n} - \frac{\frac{1}{\tau^2}}{n+\frac{1}{\tau^2}}\theta\right| > \varepsilon \middle| \Theta = \theta\right]$$

$$= \mathrm{P}\left[\left|\frac{n}{n+\frac{1}{\tau^2}}\frac{Z}{\sqrt{n}} - \frac{\frac{1}{\tau^2}}{n+\frac{1}{\tau^2}}\theta\right| > \varepsilon\right].$$

Hence $\lim_{\tau\to\infty} \mathrm{P}\left[|\delta(X)-\theta| > \varepsilon \mid \Theta = \theta\right] = \mathrm{P}[|Z| > \sqrt{n}\varepsilon] = 2\Phi\left(-\sqrt{n}\varepsilon\right)$. As $\mathrm{P}\left[|\delta(X)-\theta| > \varepsilon \mid \Theta = \theta\right]$ is uniformly bounded by 1, the dominated convergence theorem implies $\lim_{\tau\to\infty} r(\tau) = 2\Phi\left(-\sqrt{n}\varepsilon\right)$.