HW1

Gabriel ROMON

Problem 1.22

(a) The Bayes risk is defined as $\int\int L(\delta(x),\theta)f(x|\theta)\pi(\theta)dxd\theta$. Letting $m(x) = \int f(x|\theta)\pi(\theta)d\theta$ be the marginal density of X, Bayes risk can be rewritten as $\int\int L(\delta(x),\theta)\pi(\theta|x)m(x)dxd\theta$. Since the integrand is non-negative, Tonelli's theorem allows to switch the order of integration:

$$\begin{aligned}\int\int L(\delta(x),\theta)\pi(\theta|x)m(x)dxd\theta &= \int\int L(\delta(x),\theta)\pi(\theta|x)m(x)d\theta dx\\ &= \int m(x)\underbrace{\int L(\delta(x),\theta)\pi(\theta|x)d\theta}_{\text{posterior loss}}dx\end{aligned}$$

Since m is independent of δ, $\int m(x)\int L(\delta(x),\theta)\pi(\theta|x)d\theta dx$ is minimized in δ only if for each x, $\delta(x)$ minimizes $\int L(\delta(x),\theta)\pi(\theta|x)d\theta$.

(b) I assume that the norm in the statement of the problem refers to the euclidean norm in $\mathbb{R}^n$. Let x be fixed and let us look for $\delta^\pi(x) \in \arg\min_{c\in\mathbb{R}^n}\int \|h(\theta)-c\|^2\pi(\theta|x)d\theta$. Let γ be a random variable with probability density $\pi(\cdot|x)$ (that is to say γ follows the posterior distribution). Then

$$\begin{aligned}\int \|h(\theta)-c\|^2\pi(\theta|x)d\theta &= E(\|h(\gamma)-c\|^2)\\ &= E(\|h(\gamma)-E(h(\gamma))\|^2) + E(\|c-E(h(\gamma))\|^2)\\ &\quad + 2E[(h(\gamma)-E(h(\gamma)))^T(c-E(h(\gamma)))]\\ &= E(\|h(\gamma)-E(h(\gamma))\|^2) + E(\|c-E(h(\gamma))\|^2)\end{aligned}$$

which is clearly minimal for $c = E(h(\gamma))$. Hence $\delta^\pi(x) = E(h(\gamma))$, which may be rewritten with other notations as $E^\pi(h(\theta)|x)$.

(c) Let x be fixed and let us look for $\delta^\pi(x) \in \arg\min_{c\in\mathbb{R}}\int |h(\theta)-c|\pi(\theta|x)d\theta$. With the same notation as in b), $\int |h(\theta)-c|\pi(\theta|x)d\theta = E(|h(\gamma)-c|)$. We recall that a median of $h(\gamma)$ is any m such that $P(h(\gamma)\le m)\le\frac{1}{2}$ and $P(h(\gamma)\ge m)\le\frac{1}{2}$. By replacing $h(\gamma)$ with $h(\gamma)-m$ we may assume without loss of generality that $m=0$. We only need to prove that $E(|h(\gamma)-c|)\ge E(|h(\gamma)|)$.

Assume first that $c\ge 0$, and note that $\mathbb{1}_{h(\gamma)\le 0}(|h(\gamma)-c|-|h(\gamma)|) = \mathbb{1}_{h(\gamma)\le 0}(-(h(\gamma)-c)+h(\gamma)) = c$, which yields

$$E(\mathbb{1}_{h(\gamma)\le 0}[|h(\gamma)-c|-|h(\gamma)|]) = cP(h(\gamma)\le 0)$$

By the reverse triangle inequality, $|h(\gamma)-c|\ge |h(\gamma)|-|c| = |h(\gamma)|-c$, thus

$$|h(\gamma)-c|-|h(\gamma)|\ge -c$$

hence

$$E(\mathbb{1}_{h(\gamma)>0}[|h(\gamma)-c|-|h(\gamma)|]) \ge -cP(h(\gamma)>0)$$

Summing both inequalities,

$$E(|h(\gamma)-c|-|h(\gamma)|)\ge c(P(h(\gamma)\le 0)-P(h(\gamma)>0)) = c(2P(h(\gamma)\le 0)-1)\ge 0$$

since 0 is a median of $h(\gamma)$.

If $c \le 0$, simply re-use the previous case by noting that

$$E(|h(\gamma)-c|) = E(|-h(\gamma)+c|) = E(|-h(\gamma)-(-c)|) \ge E(|-h(\gamma)|) = E(|h(\gamma)|)$$

Therefore $\delta^\pi(x) = \operatorname{med} h(\gamma)$ where $\operatorname{med} h(\gamma)$ is any median of $h(\gamma)$ and γ follows the posterior distribution.

Problem 2.30

(a) Using the law of the unconscious statistician and the fact that X and U are independent,

$$\begin{aligned} P(U < \frac{f(X)}{Mg(X)}) = E(\mathbb{1}_{U<\frac{f(X)}{Mg(X)}}) &= \int\int \mathbb{1}_{u<\frac{f(x)}{Mg(x)}} \mathbb{1}_{(0,1)}(u) g(x) du dx \\ &= \int g(x) \int_0^{\min(\frac{f(x)}{Mg(x)},1)} du dx = \int g(x) \int_0^{\frac{f(x)}{Mg(x)}} du dx \quad \text{because } f \le Mg \\ &= \int g(x) \frac{f(x)}{Mg(x)} dx = \frac{1}{M} \end{aligned}$$

(b) Since f and g are normalized densities, integrating $f \le Mg$ yields $1 \le M$.

(c) The statement of the problem requires clarification. N is meant to be the number of **failed** trials until the k-th success. Let us consider $U_1, X_1, \ldots, U_{n+k}, X_{n+k}$ $n+k$ independent random variables. For $n \ge \mathbb{N}$,

$$\begin{aligned} P(N=n) &= P\left(U_{n+k} \le \frac{f(X_{n+k})}{Mg(X_{n+k})} \bigcap \sum_{i=1}^{n+k-1} \mathbb{1}_{U_i<\frac{f(X_i)}{Mg(X_i)}} = k-1\right) \\ &= P\left(U_{n+k} \le \frac{f(X_{n+k})}{Mg(X_{n+k})}\right) P\Bigg(\underbrace{\sum_{i=1}^{n+k-1} \mathbb{1}_{U_i<\frac{f(X_i)}{Mg(X_i)}}}_{\sim\mathcal{B}(n+k-1,\frac{1}{M})} = k-1\Bigg) \\ &= \frac{1}{M}\binom{n+k-1}{k-1}\frac{1}{M^{k-1}}\left(1-\frac{1}{M}\right)^n \\ &= \binom{n+k-1}{n}\frac{1}{M^k}\left(1-\frac{1}{M}\right)^n \end{aligned}$$

Hence $N \sim \mathcal{N}eg(k, \frac{1}{M})$. Thus $E(N) = \dfrac{k(1-\frac{1}{M})}{\frac{1}{M}} = k(M-1)$. The average total number of trials needed until k successes have occurred is the sum of the average number of failures and the k successes: $E(N) + k = kM$.

(d) If A verifies $f \le Ag$, then $f \le 2Ag$. To see that the bound is not tight, let $M = 2A$ and $M' = A$.

Consider for the target distribution f a Gamma(a, b) where $a \notin \mathbb{N}$. For the proposal distribution g, consider a Gamma$(\lfloor a \rfloor, \beta)$ where β must be chosen so that the domination condition holds. Note that

$$\forall x > 0, \ \frac{f(x)}{g(x)} = \frac{\Gamma(\lfloor a \rfloor)}{\Gamma(a)} \frac{b^a}{\beta^{\lfloor a \rfloor}} x^{a-\lfloor a \rfloor} e^{-(b-\beta)x}$$

the derivative of which is ≥ 0 if and only if $x \le \dfrac{a-\lfloor a \rfloor}{b-\beta}$, provided $b-\beta > 0$. Therefore, when $\beta < b$, $\frac{f}{g}$ has a maximum at $x = \frac{a-\lfloor a \rfloor}{b-\beta}$. We may now consider the choice of β that yields the tightest bound. At $\dfrac{a-\lfloor a \rfloor}{b-\beta}$, the function is proportional to $\dfrac{1}{\beta^{\lfloor a \rfloor}(b-\beta)^{a-\lfloor a \rfloor}}$, thus we must maximize $\beta^{\lfloor a \rfloor}(b-\beta)^{a-\lfloor a \rfloor}$.

A quick study of the derivative yields the optimal β as $\dfrac{\lfloor a \rfloor}{a} b$, and plugging this back into the upper bound yields the optimal bound

$$M' := \frac{\Gamma(\lfloor a \rfloor)}{\Gamma(a)} \left(\frac{a}{e}\right)^a \left(\frac{e}{\lfloor a \rfloor}\right)^{\lfloor a \rfloor}$$

However, this bound may be not computationally tractable as it requires computing special functions. For $a > 1.47$, $\frac{\Gamma(\lfloor a \rfloor)}{\Gamma(a)} \leq 1$, which yields the simpler upper bound

$$M := \left(\frac{a}{e}\right)^a \left(\frac{e}{\lfloor a \rfloor}\right)^{\lfloor a \rfloor}$$

Note that the computations above **also solve questions c) and d) in in Problem 2.32**.

(e) Let B be a measurable set and note that

$$P(X \in B | U \leq \frac{f(X)}{Mg(X)}) = \frac{P(X \in B \cap U \leq \frac{f(X)}{Mg(X)})}{P(U \leq \frac{f(X)}{Mg(X)})} = \frac{\int\int \mathbb{1}_{x \in B} \mathbb{1}_{u \leq \frac{f(x)}{Mg(x)}} \mathbb{1}_{(0,1)}(u) g(x) du dx}{\int\int \mathbb{1}_{u \leq \frac{f(x)}{Mg(x)}} \mathbb{1}_{(0,1)}(u) g(x) du dx}$$
$$= \frac{\int_B g(x) \int_0^{\min(\frac{f(x)}{Mg(x)},1)} du dx}{\int g(x) \int_0^{\min(\frac{f(x)}{Mg(x)},1)} du dx}$$

With $A = \{x, f(x) > Mg(x)\}$,

$$\begin{aligned}
P(X \in B | U \leq \frac{f(X)}{Mg(X)}) &= \frac{\int_B g(x) \int_0^{\min(\frac{f(x)}{Mg(x)},1)} du dx}{\int g(x) \int_0^{\min(\frac{f(x)}{Mg(x)},1)} du dx} \\
&= \frac{\int_B \mathbb{1}_{A^c}(x) g(x) \int_0^{\frac{f(x)}{Mg(x)}} du dx + \int_B \mathbb{1}_A(x) g(x) \int_0^1 du dx}{\int \mathbb{1}_{A^c}(x) g(x) \int_0^{\frac{f(x)}{Mg(x)}} du dx + \int \mathbb{1}_A(x) g(x) \int_0^1 du dx} \\
&= \frac{\int_B \mathbb{1}_{A^c}(x) \frac{f(x)}{M} dx + \int_B \mathbb{1}_A(x) g(x) dx}{\int \mathbb{1}_{A^c}(x) \frac{f(x)}{M} dx + \int \mathbb{1}_A(x) g(x) dx} \\
&= \int_B \frac{1}{C} \left(\mathbb{1}_{A^c}(x) f(x) + \mathbb{1}_A(x) Mg(x)\right) dx
\end{aligned}$$

Hence Y has density $\frac{1}{C}\left(\mathbb{1}_{A^c}(y) f(y) + \mathbb{1}_A(y) Mg(y)\right)$ where $C = \int \mathbb{1}_{A^c}(x) f(x) + \mathbb{1}_A(x) Mg(x) dx$. By the definition of A, this rewrites as $\frac{1}{C} \min(f(y), Mg(y))$ which is different from $f(y)$ in general. Therefore, when the bound is too tight, the algorithm does not sample according to f.

(f) Suppose that $\forall x, f(x) \leq M' g(x) < Mg(x)$. Let us determine the distribution of $X | U > \frac{f(X)}{Mg(X)}$. Let B be a measurable set.

$$\begin{aligned}
P(X \in B | U > \frac{f(X)}{Mg(X)}) &= \frac{\int\int \mathbb{1}_{x \in B} \mathbb{1}_{u > \frac{f(x)}{Mg(x)}} \mathbb{1}_{(0,1)}(u) g(x) du dx}{\int\int \mathbb{1}_{u > \frac{f(x)}{Mg(x)}} \mathbb{1}_{(0,1)}(u) g(x) du dx} \\
&= \frac{\int\int \mathbb{1}_{x \in B} g(x) \int_{\frac{f(x)}{Mg(x)}}^1 du dx}{\int\int g(x) \int_{\frac{f(x)}{Mg(x)}}^1 du dx} \\
&= \frac{\int_B g(x) - \frac{1}{M} f(x) dx}{1 - \frac{1}{M}}
\end{aligned}$$

Let Z be the random variable that samples failed trials. Considering the previous computations, Z has density $\dfrac{g(z) - \frac{1}{M} f(z)}{1 - \frac{1}{M}}$. This density may be used as a proposal distribution since

$$\frac{f(z)}{\frac{g(z) - \frac{1}{M} f(z)}{1 - \frac{1}{M}}} \leq \frac{f(z)}{\frac{\frac{1}{M'} f(z) - \frac{1}{M} f(z)}{1 - \frac{1}{M}}} = \frac{M - 1}{\frac{M}{M'} - 1}$$

Problem 3.22

(a) The expectation of the sum of the weights is $E\left(\sum_{j=1}^{n} \frac{f(X_j)}{g(X_j)}\right) = \sum_{j=1}^{n} E\left(\frac{f(X_j)}{g(X_j)}\right)$. Since the X_j are sampled according to g, $E\left(\frac{f(X_j)}{g(X_j)}\right) = \int g(x)\frac{f(x)}{g(x)}dx = 1$. Hence $E\left(\sum_{j=1}^{n} \frac{f(X_j)}{g(X_j)}\right) = n$.

However, there is no reason for $\sum_{j=1}^{n} \frac{f(X_j)}{g(X_j)} = n$ to hold in general. If the weights are to be used for resampling, one must normalize them properly as $\frac{w_i}{\sum_{i=1}^{n} w_i}$.

(b) We assume that the weights have been normalized to sum to 1. For each j, $\tilde{X}_j$ is defined by its conditional distribution: $\tilde{X}_j | X_1, \ldots, X_n \sim \sum_{i=1}^{n} w_i \delta_{X_i}$. Thus

$$E(h(\tilde{X}_j)) = E(E(h(\tilde{X}_j)|X_1, \ldots, X_n)) = E(\sum_{i=1}^{n} w_i h(X_i))$$

Hence

$$E\left(\frac{1}{n}\sum_{j=1}^{n} h(\tilde{X}_j)\right) = E\left(\sum_{i=1}^{n} w_i h(X_i)\right)$$

(c) If the formula holds with $w_i = \frac{1}{n}\frac{f(X_i)}{g(X_i)}$, then

$$E\left(\frac{1}{n}\sum_{j=1}^{n} h(\tilde{X}_j)\right) = E\left(\sum_{i=1}^{n} \frac{1}{n}\frac{f(X_i)}{g(X_i)} h(X_i)\right) = n \cdot \frac{1}{n} \int g(x)\frac{f(x)}{g(x)}h(x)dx = \int h(x)f(x)dx$$

With $h : x \mapsto \mathbb{1}_{x \le t}$, one gets

$$E\left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\tilde{X}_j \le t}\right) = \int_{-\infty}^{t} f(x)dx$$

Therefore, the cumulative distribution of the $\tilde{X}_j$ is unbiased.