Chapter 2

Lecture 6

Bayes estimation

We have the setup from the previous lecture: $(S, \mathcal{A}, P_\theta)$, $\theta \in \Theta$. We want to estimate $g(\theta)$. Let $\mathcal{B}$ be a σ-field in Θ and λ a probability on $\mathcal{B}$. We assume that g is $\mathcal{B}$-measurable and that $g \in L^2(\Theta, \mathcal{B}, \lambda)$ – i.e., that $\int_\Theta g(\theta)^2 d\lambda(\theta) < \infty$. We regard θ as a random element and $P_\theta(A)$ as a conditional probability that $s \in A$, given θ. Let $w = (s, \theta)$, $\Omega = S \times \Theta$ and $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ be the smallest σ-field containing all sets of the form $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. We assume that $\theta \mapsto P_\theta(A)$ is $\mathcal{B}$-measurable for all $A \in \mathcal{A}$.

Lemma 1. *There exists a unique probability measure M on $\mathcal{C}$ such that*

$$M(A \times B) = \int_B P_\theta(A) d\lambda(\theta) \ \forall A \in \mathcal{A}, B \in \mathcal{B}.$$

(λ is the distribution of θ, P_θ is the distribution of s given θ and M is the joint distribution of $w = (s, \theta)$. We will see the explicit formula for Q_s – the distribution of θ given s – soon.)

Consider an estimate t. Our assumption on g is that

$$E(g^2) = \int_\Omega g^2 dM = \int_\Theta g^2 d\lambda < \infty,$$

so

$$\begin{aligned} \overline{R}_t = \int_\Theta R_t(\theta) d\lambda(\theta) = \int_\Theta E_\theta\big(t - g(\theta)\big)^2 d\lambda(\theta) \\ = E\big(E((t-g)^2 \mid \theta)\big) = E(t-g)^2 = ||t-g||^2 \end{aligned}$$

(the norm taken in $L^2(\Omega, \mathcal{C}, M)$). We would like to choose a t to minimize this quantity. Since t is a function only of s, the desired minimizing estimate – which we will denote by t^* – is the projection of g to the subspace of all $\mathcal{A}$-measurable functions $t(s)$ satisfying $E_M\big(t(s)^2\big) < \infty$. We know that $t^*(s) = E\big(g(\theta) \mid s\big)$.

4. a. There exists a minimizing t^* (this is the Bayes estimate for g with respect to λ).

   b. The minimizing t^* is unique up to $\overline{P}$-equivalence, where

$$\overline{P}(A) = \int_\Theta P_\theta(A)d\lambda(\theta)$$

   for $A \in \mathcal{A}$.

   c.

$$\begin{aligned}\overline{R}_{t^*} &= \inf_t \overline{R}_t = ||t^* - g||^2 = ||g||^2 - ||t^*||^2 \\ &= \int_\Omega g^2 dM - \int_\Omega (t^*)^2 dM = \int_\Theta g(\theta)^2 d\lambda(\theta) - \int_S t^*(s)^2 d\overline{P}(s).\end{aligned}$$

   d. $t^*(s) = E\big(g(\theta) \mid s\big)$, where (s, θ) is distributed according to M.

   Proof. Clear. □

Note. $\overline{P}$ is the marginal distribution of s.

Some explicit formulas

Suppose we begin with a dominated family $\{P_\theta : \theta \in \Theta\}$ – i.e., a family such that there exists a σ-finite μ such that each P_θ admits a density, say ℓ_θ with respect to μ such that ℓ_θ is measurable, $0 \leq \ell_\theta(s) < \infty$ and

$$P_\theta(A) = \int_A \ell_\theta(s)d\mu(s)\ \forall A \in \mathcal{A}.$$

Two basic examples are:

i. $S = \{s_1, s_2, \ldots\}$ is countable and μ is counting measure. Then $\ell_\theta(s)$ is the P_θ-probability of the atomic set $\{s\}$.

ii. $S = \mathbb{R}^k$, $\mathcal{A} = \mathcal{B}^k$ and μ is Lebesgue measure on $\mathbb{R}^k$. Then ℓ_θ is the probability density in the familiar sense.

We assume that $(s, \theta) \mapsto \ell_\theta(s)$ is a $\mathcal{C}$-measurable function. We write $\nu = \mu \times \lambda$.

5. a. M is given by

$$\frac{dM}{d\nu}(s, \theta) = \ell_\theta(s).$$

   b. $\overline{P}$ is given by

$$\frac{d\overline{P}}{d\mu}(s) = \int_\Theta \ell_\theta(s)d\lambda(\theta).$$

c. For all $s \in S$, Q_s, the conditional probability on Θ given s when (s, θ) is distributed according to M, is well-defined and given by

$$\frac{dQ_s}{d\lambda}(\theta) = \frac{\ell_\theta(s)}{\overline{\ell}(s)},$$

where $\overline{\ell}(s) = \frac{d\overline{P}}{d\mu}$.

d. For all $g \in L^2$,

$$E(g \mid s) = t^*(s) = \int_\Theta g(\theta) dQ_s(\theta) = \int_\Theta g(\theta) \frac{\ell_\theta(s)}{\overline{\ell}(s)} d\lambda(\theta).$$

Proof. These are all easy consequences of Fubini's theorem. (For example, since

$$\begin{aligned}
\Pr(s \in A, \theta \in B) = E\big(I_A(s) I_B(\theta)\big) &= E\big(I_B(\theta) E(I_A(s) \mid \theta)\big) \\
&= E\big(I_B(\theta) P_\theta(A)\big)
\end{aligned}$$

$$= E\Big(I_B(\theta) \int_A \ell_\theta d\mu\Big) = \int_\Theta I_B(\theta) \Big[\int_A \ell_\theta(s) d\mu(s)\Big] d\lambda(\theta) = \int_{A \times B} \ell_\theta(s) d\nu(s, \theta),$$

we have (a).) □

Lecture 7

Note. In any $V = L^2(\Omega, \mathcal{C}, M)$, the constant functions form a subspace, which we will denote by $W_c = W_c(\Omega, \mathcal{C}, M)$. The projection of $f \in V$ on W_c is just $E(f)$.

Note. In the present context, with $w = (s, \theta)$ and $\Omega = S \times \Theta$, s is the datum and θ is the unknown parameter. λ is the prior distribution of θ, Q_s is the posterior distribution (after s is observed) of θ, $t^*(s) = E(g \mid s)$ is the posterior mean of $g(\theta)$, $\theta \mapsto \ell_\theta(s)$ is the likelihood function and $\ell_\theta = \frac{dP_\theta}{d\mu}$.

Note. If we do have a λ on hand but no data, then the Bayes estimate is just $Eg = \int_\Theta g(\theta) d\lambda(\theta)$.

6. If t^* is a Bayes estimate of g, then t^* cannot be unbiased, except in trivial cases.

   Proof. Suppose that t^* is unbiased and Bayes (with respect to λ). Then, by unbiasedness,

   $$(t^*, g) = E(t^* g) = E\big(E(t^* g \mid \theta)\big) = E\big(g(\theta) E(t^* \mid \theta)\big) = E(g(\theta)^2) = ||g||^2;$$

   but also, since t^* is a Bayes estimate,

   $$(t^*, g) = E\big(E(t^* g \mid s)\big) = E\big(t^*(s) E(g \mid s)\big) = ||t^*||^2.$$

From this we conclude that:

$$\begin{aligned}||t^* - g||^2 = ||g||^2 - ||t^*||^2 = 0 &\Leftrightarrow M(\{w : t^*(s) \neq g(\theta)\}) = 0 \\ \Leftrightarrow \int_\Theta P_\theta(t^*(s) \neq g(\theta))d\lambda(\theta) = 0 &\Leftrightarrow P_\theta(t^*(s) = g(\theta)) = 1 \text{ a.e.}(\lambda)\end{aligned}$$

This last statement, though, holds iff there is an estimate t such that $P_\theta(t(s) = g(\theta)) = 1$ a.e.(λ). Hence, except in the trivial case, t^* cannot be both Bayes and unbiased. $\square$

Example 1(a). $s = (X_1, X_2, \dots, X_n)$, $X_i \overset{\text{iid}}{\sim} N(\theta, 1)$ and $\Theta = \mathbb{R}^1$. Let μ be Lebesgue measure; then $d\mu = dX_1 \cdots dX_n$ and $\ell_\theta(s) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i - \theta)^2}$. Consider the estimation of $g(\theta) = \theta$. $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is an unbiased estimate of g (in fact, it is the UMVUE of g), so $\overline{X}$ is not Bayes with respect to any λ. We will see later that:

i. $\overline{X}$ is minimax,

ii. $\overline{X}$ is the pointwise limit of Bayes estimates and

iii. $\overline{X}$ is admissible.

Suppose that λ is the $N(0, \sigma^2)$ distribution, so that $d\lambda(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\theta^2} d\theta$. Given s, $\frac{dQ_s}{d\lambda}(\theta)$ is proportional to $\ell_\theta(s)$, i.e., $dQ_s(\theta) = \varphi_1(s)\ell_\theta(s)d\lambda(\theta)$ for some function φ_1 of s.

Now

$$\ell_\theta(s) = \left(\frac{1}{2\pi}\right)^n e^{-\frac{n}{2}(\overline{X}-\theta)^2 - \frac{n}{2}v},$$

where $v = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$, so that $dQ_s(\theta) = \varphi_1(s)\varphi_2(s)e^{-\frac{n}{2}(\overline{X}-\theta)^2} e^{-\frac{\theta^2}{2\sigma^2}} d\theta$. Let $\sigma^2 = \frac{1}{nh^2}$ for some $h > 0$, so that

$$dQ_s(\theta) = \varphi_3(s)e^{n\overline{X}\theta - \frac{n}{2}h^2\theta^2 - \frac{n}{2}\theta^2} d\theta = \varphi_4(s)e^{-\frac{n(1+h^2)}{2}\left[\theta - \frac{\overline{X}}{1+h^2}\right]^2} d\theta.$$

Thus $Q_s \sim N\left(\frac{\overline{X}}{1+h^2}, \frac{1}{n(1+h^2)}\right)$ (where $h^2 = \left(\frac{1}{n}\right)/\sigma^2$ is the ratio of observation variance to prior variance) and $t^*(s) = E(\theta \mid s) = \frac{\overline{X}}{1+h^2}$. Therefore

$$t^*(s) = \frac{h^2}{1+h^2} \cdot \underbrace{0}_{\dagger} + \frac{1}{1+h^2}\overline{X} = \underbrace{\frac{1/\sigma^2}{n + 1/\sigma^2} \cdot 0 + \frac{n}{n + 1/\sigma^2}\overline{X}}_{\ddagger},$$

where the 0 (labelled † above) arises from the fact that we are dealing with a Bayes estimate with no data; and the terms marked ‡ represent a weighted average of prior and data mean, with weights proportional to inverse variance.

We have

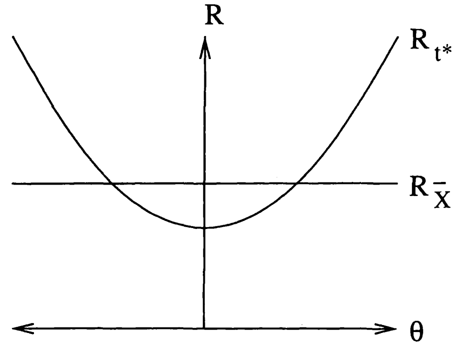
$$R_{\overline{X}}(\theta) = \text{Var}_\theta(\overline{X}) + 0 = \frac{1}{n},$$

$$R_{t^*}(\theta) = \mathrm{Var}_\theta(t^*) + \left(\frac{\theta}{1+h^2} - \theta\right)^2 = \frac{1}{(1+h^2)^2}\frac{1}{n} + \left(\frac{h^2}{1+h^2}\right)^2 \theta^2$$

and

$$\overline{R}_{t^*} = \frac{1}{(1+h^2)^2}\frac{1}{n} + \left(\frac{h^2}{1+h^2}\right)^2 \sigma^2 = \frac{1}{n(1+h^2)}.$$

(The last can be checked by noting that $\overline{R}_{t^*} = ||\theta - t^*||^2 = ||\theta||^2 - ||t^*||^2$.)



Lecture 8

In the framework $(S, \mathcal{A}, P_\theta)$, $\theta \in \Theta$, set up in the previous lectures, suppose that $dP_\theta = \ell_\theta d\mu$ on S, where $d\mu$ is a fixed measure (this is another way of saying that ℓ_θ is the likelihood function). Given s, suppose $\hat{\theta} = \hat{\theta}(s)$ is the point in Θ such that $\ell_{\hat{\theta}(s)}(s) = \sup_{\theta\in\Theta} \ell_\theta(s)$. Then $\hat{\theta}$ is the (or an) ML estimate of θ.

Given a function g on Θ, $g(\hat{\theta}(s))$ is the ML estimate of $g(\theta)$.

Homework 2

1. What is $\hat{\theta}$ in Example 1(b) (on page 10)? In Example 1(c) (no explicit answer is available in this case)? Assume that $\Theta = \mathbb{R}^1$.

We return to our investigation of Example 1(a).

Example 1. Here μ is n-dimensional Lebesgue measure and $\ell_\theta(s) = \varphi(s)e^{-\frac{n}{2}(\overline{X}-\theta)^2}$, so that $\hat{\theta}(s) = \overline{X}$. The ML estimate of θ^2 is $\overline{X}^2$, the ML estimate of $|\theta|$ is $|\overline{X}|$, etc.

Under square error loss, we know that, if λ is distributed as $N(0, \sigma^2)$ with $\sigma^2 = \frac{1}{nh^2}$, then the Bayes estimate of θ is $t^*(s) = \frac{\overline{X}}{1+h^2}$. The Bayes estimate of θ^2 is

$$\int_{\mathbb{R}^1} \theta^2 dQ_s(\theta) = \frac{1}{n(1+h^2)} + \frac{\overline{X}^2}{(1+h^2)^2}$$

and the Bayes estimate of $|\theta|$ is $\int_{\mathbb{R}^1} |\theta| dQ_s(\theta)$, which is a sum involving special functions.

Note.

i. We have seen that the Bayes estimates of θ converge to $\overline{X}$, which is ML and UMVUE, as $h \to 0$ (i.e., $\sigma^2 \to +\infty$). This does not always happen, however: The Bayes estimates of θ^2 converge to $\overline{X}^2 + \frac{1}{n}$; the ML estimate of θ^2 is $\overline{X}^2$ and the UMVUE of θ^2 is $\overline{X}^2 - \frac{1}{n}$. These three are not identical, but they are close if n is large.

ii. The case $h = 0$ ($\sigma^2 = +\infty$) corresponds to uniform prior ignorance about θ; i.e., for two intervals (a, b) and (c, d) in $\mathbb{R}^1 = \Theta$, $\frac{\lambda(a,b)}{\lambda(c,d)} \to \frac{b-a}{d-c}$ as $h \downarrow 0$ (from Homework 2).

iii. $t^*(s) = \frac{\overline{X}}{1+h^2}$ (h fixed) is admissible because t^* is an essentially unique Bayes estimate in $L^2(M)$ (i.e., if t_0 is also Bayes, then $\overline{P}(t_0 = t^*) = 1$, which is equivalent to saying that $P_\theta(t_0 = t^*) = 1$ for all $\theta \in \Theta$). It follows that, for any constant c, $t(s) = \alpha\overline{X} + (1-\alpha)c$ is admissible for any $0 \le \alpha \le 1$. (Let $X_i' = X_i + c$ and $\theta' = \theta + c$, and apply the above result.)

Homework 2

2. Find a necessary and sufficient condition such that $b_0 + b_1X_1 + \cdots + b_nX_n$ be admissible for θ.

For any $h > 0$, $R_{t^*}(\theta) = u + v\theta^2$, so $\sup_{\theta\in\Theta} R_{t^*}(\theta) = +\infty$ and t^* is not minimax. We do have, however, that:

i. $\overline{X}$ is minimax.

Proof. Choose any estimate t.

$$\sup_\theta R_t(\theta) \ge \overline{R}_t \ge \overline{R}_{t^*} = \frac{1}{n(1+h^2)};$$

so, since h is arbitrary, $\sup_\theta R_t(\theta) \ge \frac{1}{n}$, which is the (constant) risk of $\overline{X}$. Thus $\overline{X}$ is minimax. □

ii. $\overline{X}$ is admissible.

Proof. Let t be an estimate such that

$$R_t(\theta) \le R_{\overline{X}}(\theta) = \frac{1}{n}\ \forall\theta \in \Theta$$

and let, for $h > 0$, λ be distributed as $N(0, 1/nh^2)$. We have that $\overline{R}_t \ge \overline{R}_{t^*}$ because t^* is Bayes for λ.

Now $t-\theta = (t-t^*) + (t^*-\theta)$ and $(t-t^*) \perp (t^*-\theta)$, so

$$||t-t^*||^2 = ||t-\theta||^2 - ||t^*-\theta||^2 = \overline{R}_t - \overline{R}_{t^*}$$
$$\leq \frac{1}{n} - \overline{R}_{t^*} = \frac{1}{n} - \frac{1}{n(1+h^2)} = \frac{h^2}{n(1+h^2)}.$$

We have also that

$$\begin{aligned} ||t-t^*||^2 &= \int_{\mathbb{R}^{n+1}} (t-t^*)^2 dM = \int_{\mathbb{R}^n} (t-t^*)^2 d\overline{P} \\ &= \int_{\mathbb{R}^n} \big(t(s)-t^*(s)\big)^2 \int_\Theta \frac{\sqrt{n}h}{\sqrt{2\pi}} e^{-\frac{nh^2}{2}\theta^2} \ell_\theta(s) d\theta\, d\mu(s) \\ &= \int_{\mathbb{R}^{n+1}} \big(t(s)-t^*(s)\big)^2 \frac{\sqrt{n}h}{\sqrt{2\pi}} e^{-\frac{nh^2}{2}\theta^2} \ell_\theta(s) d\theta\, d\mu(s). \end{aligned}$$

From these two equalities we conclude that

$$\int_{\mathbb{R}^{n+1}} \big(t(s)-t^*(s)\big)^2 \sqrt{\frac{n}{2\pi}} e^{-\frac{nh^2}{2}\theta^2} \ell_\theta(s) d\mu(s) d\theta \leq \frac{h}{n(1+h^2)}.$$

Letting $h \to 0$ and using Fatou's lemma and the fact that $t^* \to \overline{X}$ (as $h \to 0$), we have that

$$\int_{\mathbb{R}^{n+1}} \big(t(s)-\overline{X}\big)^2 \ell_\theta(s) d\mu(s) d\theta = 0$$

and hence that $\big(t(s)-\overline{X}\big)^2 \ell_\theta(s) = 0$ a.e. (with respect to Lebesgue measure) in $\mathbb{R}^{n+1}$. Since $\ell_\theta(s) > 0$ for all $(s,\theta) \in \mathbb{R}^{n+1}$ by presumption, we have that

$$\begin{aligned} &(t(s)-\overline{X} = 0 \text{ a.e. (with respect to Lebesgue measure) on } \mathbb{R}^{n+1}) \\ &\quad \Rightarrow (t(s) = \overline{X} \text{ a.e. (with respect to Lebesgue measure) on } S = \mathbb{R}^n) \\ &\quad\quad \Rightarrow (P_\theta\big(t(s) = \overline{X}\big) = 1\ \forall\theta \in \Theta) \Rightarrow (R_t(\theta) = R_{\overline{X}}(\theta) = \frac{1}{n}\ \forall \theta \in \Theta). \end{aligned}$$

□

Lecture 9

Homework 2

3. In Example 1(a), what is the marginal distribution $\overline{P}$ of s? Are the X_is normal and independent under $\overline{P}$? What is the distribution of $\overline{X}$ under $\overline{P}$? Here the prior is assumed to be $\lambda \sim N(0, 1/nh^2)$.

Example 2(a). Let n be a fixed positive integer and $s = (X_1, \ldots, X_n)$ with the X_i iid random variables assuming the values 1 and 0 with probabilities θ and $1 - \theta$ respectively and $\Theta = [0, 1]$. Let μ be counting measure on S, the set of 2^n possible values assumed by s. Then

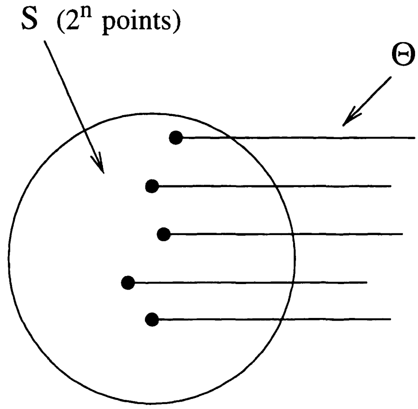
$$\ell_\theta(s) = P_\theta(\{s\}) = \theta^{T(s)}(1-\theta)^{n-T(s)}$$

where $T(s) = \sum_{i=1}^n X_i$. The ML estimate is $\hat{\theta}(s) = T(s)/n = \overline{X}$, which is unbiased for θ. (We will see that in fact $\overline{X}$ is the UMVUE.) $R_{\overline{X}}(\theta) = \mathrm{Var}_\theta(\overline{X}) = \theta(1-\theta)/n$. The ML estimate of $\theta(1-\theta)/n$ is just $\overline{X}(1-\overline{X})/n$, which is *not* unbiased. We shall see later that $\frac{1}{n}\left(\frac{T}{n} - \frac{T}{n}\frac{T-1}{n-1}\right)$ is the UMVUE for $\theta(1-\theta)/n$. Let λ be the Beta distribution $B(a,b)$, i.e.,

$$d\lambda(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta$$

for $0 \le \theta \le 1$, with parameters $a, b > 0$. Here $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the Beta function; it is easy to check that $E_\lambda(\theta) = \frac{a}{a+b}$ and $\mathrm{Var}_\lambda(\theta) = \frac{ab}{(a+b)^2(a+b+1)}$.

We visualize the product space $\Omega = S \times \Theta$ as a unit interval attached to each point of S:



We let μ be counting measure on S, λ prior measure on Θ and $\nu = \mu \times \lambda$ the product measure on Ω. We have $dM(s,\theta) = \ell_\theta(s) d\nu(s,\theta)$ on Ω, so that $M(C) = \int_C \ell_\theta(s) d\nu(s,\theta)$ for all $C \in \mathcal{C} = \mathcal{A} \times \mathcal{B}$, and

$$\begin{aligned} dQ_s(\theta) = \frac{\ell_\theta(s)}{\bar{\ell}(s)} d\lambda(\theta) &= \varphi_1(s)\ell_\theta(s) d\lambda(\theta) \\ &= \varphi_2(s)\theta^{T(s)}(1-\theta)^{n-T(s)} d\lambda(\theta) \\ &= \varphi_2'(s)\theta^{a+T(s)-1}(1-\theta)^{b+n-T(s)-1} d\theta, \end{aligned}$$

where $\varphi_2'(s) = B(a+T(s), b+n-T(s))^{-1}$, so that Q_s is the $B(a+T(s), b+n-T(s))$ distribution.

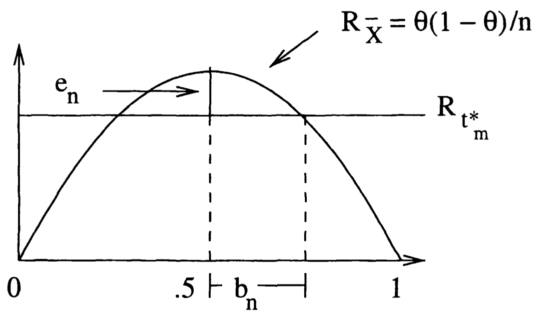
The Bayes estimate for θ is $t^*(s) = E_{Q_s}(\theta) = \frac{a+T(s)}{a+b+n}$; similarly, the Bayes estimate for $\theta(1-\theta)$ is $E_{Q_s}(\theta) - E_{Q_s}(\theta^2)$.

$$R_{t^*}(\theta) = \mathrm{Var}(t^*) + \left[b_{t^*}(\theta)\right]^2 = \frac{n\theta(1-\theta)}{(a+b+n)^2} + \left[\frac{a+n\theta}{a+b+n} - \theta\right]^2.$$

If we choose $a = \frac{\sqrt{n}}{2} = b$, then t^* becomes $t^*_m = \frac{T+\sqrt{n}/2}{n+\sqrt{n}}$ and

$$R_{t^*_m}(\theta) = \frac{n}{4(\sqrt{n}+n)^2} < \frac{1}{4n}.$$

Hence t^*_m is a Bayes estimate with constant risk; therefore it must be minimax. The graphs for the risk functions of $\overline{X}$ and t^*_m look like:



As $n \to \infty$, $b_n \to 0$ and $\frac{e_n}{1/(4n)} \to 0$. Neither $\overline{X}$ nor t^*_m is perfect; for example, if $n = 100$ and $T = 0$, then $\overline{X} = \hat{\theta} = 0$, which is too low, but $t^*_m = \frac{0+5}{100+10} \approx 4\frac{1}{2}\%$, which may be too high.

Note. With $a = \frac{\sqrt{n}}{2} = b$, the prior mean is $\frac{1}{2}$ and the prior variance is $\frac{n/4}{n(\sqrt{n}+1)} = \frac{1}{4(\sqrt{n}+1)}$.

Homework 2

4. Show that $\overline{X} = \frac{T(s)}{n}$ is admissible in two ways:

   a. Show that $\overline{X}$ is the pointwise limit of t^* as $a \downarrow 0$ and $b \downarrow 0$.

   b. Redefine the loss function by $L(t,\theta) = \frac{(t-\theta)^2}{\theta(1-\theta)}$, so that $R_t(\theta) = \frac{E_\theta(t-\theta)^2}{\theta(1-\theta)}$. (Admissibility with respect to this loss function is equivalent to admissibility with respect to the loss function $L(t,\theta) = (t-\theta)^2$.)

5. Show that $\overline{X}$ is the unique Bayes estimate with respect to some λ.

Lecture 10

Example 2(b) (negative binomial sampling). We have $\theta \in \Theta = (0,1)$. We choose a positive integer k and observe the iid

$$X_i = \begin{cases} 1 & \text{with probability } \theta \\ 0 & \text{with probability } 1-\theta \end{cases}$$

until exactly k 1's are observed.

Let N be the total number of X_is observed. Here $s = (X_1, \ldots, X_N)$ and N is a random variable.

Now let S be the set of all possible values of s; then S is countable (obviously $N \geq k$. There are $\binom{r-1}{k-1}$ values of s corresponding to $N = r$). Let μ be counting measure on S (such that sample points with N the same have the same probability). Then

$$\ell_\theta(s) = P_\theta(\text{observing } s) = \theta^{k-1}(1-\theta)^{(N(s)-1)-(k-1)}\theta = \theta^k(1-\theta)^{N(s)-k}.$$

The MLE of θ is $\overline{X} = \frac{k}{N(s)}$, which is not unbiased. Note that $N = N_1 + \cdots + N_k$, where N_1 is the number of trials until the first 'success' (i.e., observation of a 1), N_2 is the number of additional trials required for the second success etc. The N_is are iid, so that $E_\theta(N) = kE_\theta(N_1)$ and $\operatorname{Var}_\theta(N) = k\operatorname{Var}_\theta(N_1)$. Since $P_\theta(N_1 = r) = (1-\theta)^{r-1}\theta$ (for $r = 1, 2, \ldots$), we have $E_\theta(N_1) = 1/\theta$ and $\operatorname{Var}_\theta(N_1) = (1-\theta)/\theta^2$. Thus

$$E_\theta\left(\frac{N}{k}\right) = \frac{1}{k}E_\theta(N) = E_\theta(N_1) = \frac{1}{\theta};$$

remember, however, that, by the Cauchy-Schwarz inequality, $E(X)E(1/X) \geq 1$ for any random variable $X > 0$, with equality iff $P(X = c) = 1$, and so

$$E_\theta\left(\frac{k}{N}\right) > \frac{1}{1/\theta} = \theta$$

– i.e., $\overline{X} = \frac{k}{N(s)}$ is biased upwards.

It can be shown by the Rao-Blackwell theorem that the estimate $t = \frac{k-1}{N-1}$ is unbiased when $k \geq 2$. In fact, t is (by the Lehmann-Scheffé theorem or a geometrical approach) the UMVUE. (Heuristically, we see that, if $s = (X_1, \ldots, X_{N-1}, X_N)$, then necessarily $X_N \equiv 1$ (we stop as soon as we observe the kth 1) and so only $(X_1, \ldots, X_{N-1})$ constitute the active part. Then $t(s) = \frac{\text{number of successes in active part}}{\text{number of trials in active part}}$.) To see that t is unbiased, note that

$$P_\theta(N = r) = \binom{r-1}{k-1}\theta^k(1-\theta)^{r-k}$$

for $r = k, k+1, \ldots$, so that

$$E_\theta(t) = \sum_{r=k}^{\infty} \frac{k-1}{r-1} P_\theta(N = r) = \theta.$$

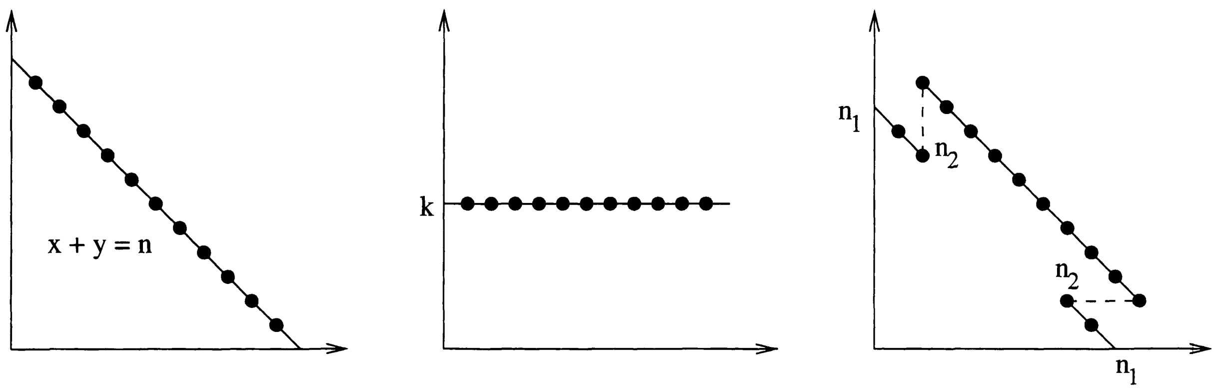
We have $\overline{X} - t(s) = \frac{k}{N(s)} - \frac{k-1}{N(s)-1} = \frac{N(s)-k}{N(s)(N(s)-1)} \geq 0$, the inequality being strict with positive probability; so $E_\theta(\overline{X}) > E_\theta(t) = \theta$.

Bayes estimates

Let λ be a prior probability measure on $(0,1)$. As always $dQ_s(\theta) = \varphi_1(s)\ell_\theta(s)d\lambda(\theta)$. Since $\ell_\theta(s)$ is as in Example 2(a), formally the Bayes estimate here is identical to the one there. In particular, $\frac{a+k}{a+b+N(s)}$ is admissible and Bayes with respect to the $B(a,b)$ prior with $a, b > 0$.

Note. Although the MLEs in Examples 2(a) and 2(b) are formally identical, the risk functions are different. In Example 2(b), $R_{\overline{X}}(\theta) = \mathrm{Var}_\theta(\overline{X}) + \left[E_\theta(\overline{X} - \theta)\right]^2$ and $R_t(\theta) = \mathrm{Var}_\theta(t)$ are complicated expressions.

Example 2(c). Depicted here are the stopping points for Examples 2(a) and 2(b), along with those of another possible (two-stage) sampling scheme:



Example 2(a) Example 2(b) Example 2(c)

Here (as in any scheme) the likelihood function is $\ell_\theta(s) = \theta^{T(s)}(1-\theta)^{N(s)-T(s)}$, where $T(s)$ is the number of successes (and, of course, $N(s) - T(s)$ is the number of failures). μ is counting measure and the MLE is $\hat{\theta}(s) = \frac{T(s)}{N(s)}$ always.

$s = (X_1, \ldots, X_N)$, where $N = n_1$ or $N = n_1 + n_2$ depending on s. How do we estimate θ? What is the precision of this estimate?