Stat 543 Homework 1Solution

1.

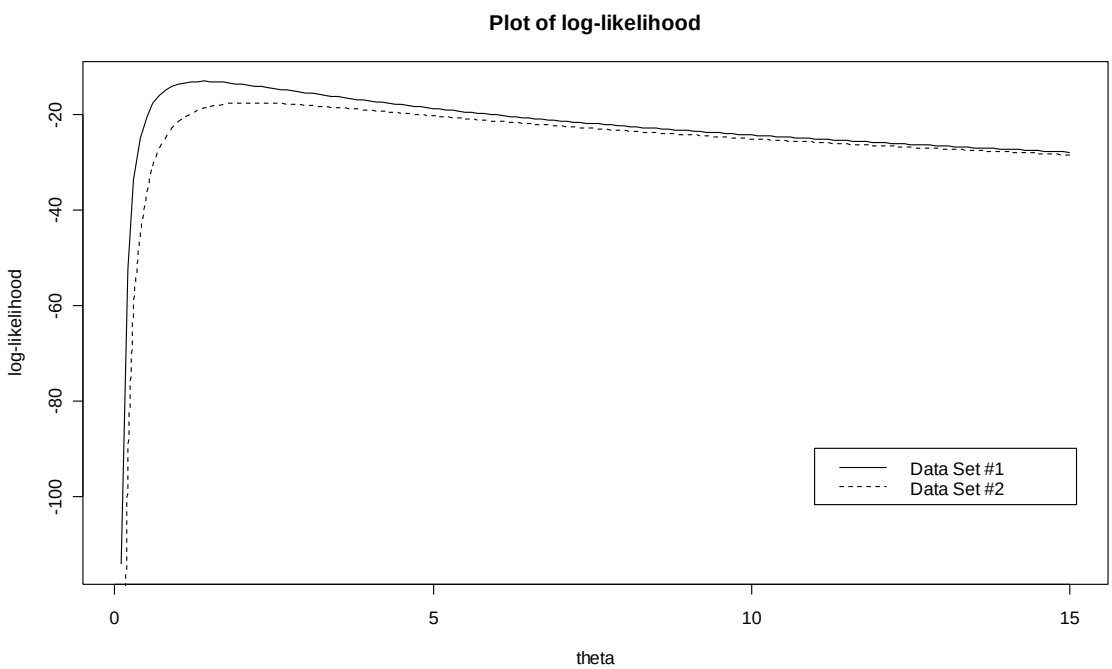
In this problem we'll consider some issues concerning the defining and meaning of a "likelihood function." Underlying what follows, suppose that $X_1, X_2, \ldots, X_{10}$ are iid exponential with mean θ. That is, we'll suppose that the marginal *pdf* is

$$f(x|\theta) = I[x \geq 0]\frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)$$

Here are two data sets (pseudo-randomly) generated from this model with $\theta = 2$.

Data Set #1 : 2.661, 3.764, 0.079, 1.286, 0.081, 0.233, 2.283, 2.600, 0.321, 0.421
Data Set #2 : 2.426, 1.852, 0.803, 11.524, 0.055, 1.266, 0.506, 0.448, 0.749, 1.930

(a) Plot, on the same set of axes, the natural logarithms of the likelihood functions (as defined in class for jointly continuous observations) based on Data Sets #1 and #2. (It is often convenient to work on a log-likelihood scale, especially in models with independent observations.) Are these two functions exactly the same? Do they provide some reliable indication of how Vardeman generated the data?

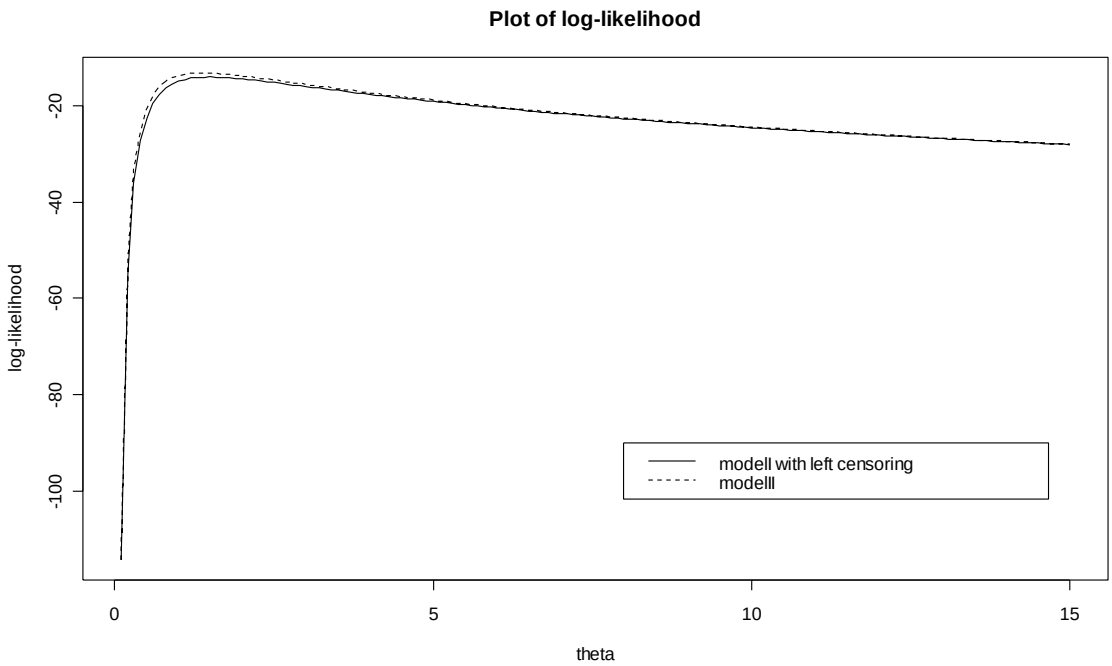


From the plot of the log-likelihood functions based on Data Set #1 and #2, we can see that both curves are almost in the same shape and reach their maximum at around $\theta = 2$, which does provide some reliable indication that both data sets are generated from the given model with $\theta = 2$.

(b) (Type I left censoring) Suppose that what is observable is not *Xi*, but rather *Wi = XiI [Xi > 1]*. That is, if an observation is less than 1, I observe a "0" (that I might as well take as a code for "observation was less than 1"). In this circumstance, a sensible definition of a likelihood function would seem to be

$$L_{\text{left}}(\theta) = \prod_{i=1}^{10} \left(I[w_i = 0] \left(1 - \exp\left(-\frac{1}{\theta}\right)\right) + I[w_i \geq 1] \frac{1}{\theta} \exp\left(-\frac{w_i}{\theta}\right) \right)$$

Plot the logarithm of this likelihood function for Data Set #1. How does it compare to that for Data Set #1 in part (a)? Does it have a similar "vertical position" on the plot? Does it have a similar "shape"?



From the plot above, we can see that the curve of the log-likelihood function of the model with left censoring is in the same shape with the one of the original model; and they almost have the same vertical position.

(c) (Rounding/digitalization) Using the joint *pdf* as a likelihood function essentially means that regardless of how many digits are written down, a data value entered to make the likelihood is treated as an "exact" "real number." That is, "2.661" implicitly has an infinite number of trailing 0's behind it. Another (more realistic) point of view is to think of "2.661" as meaning "something between 2.6605 and 2.665." Under this interpretation, a more sensible likelihood function than the one used in part (a) is

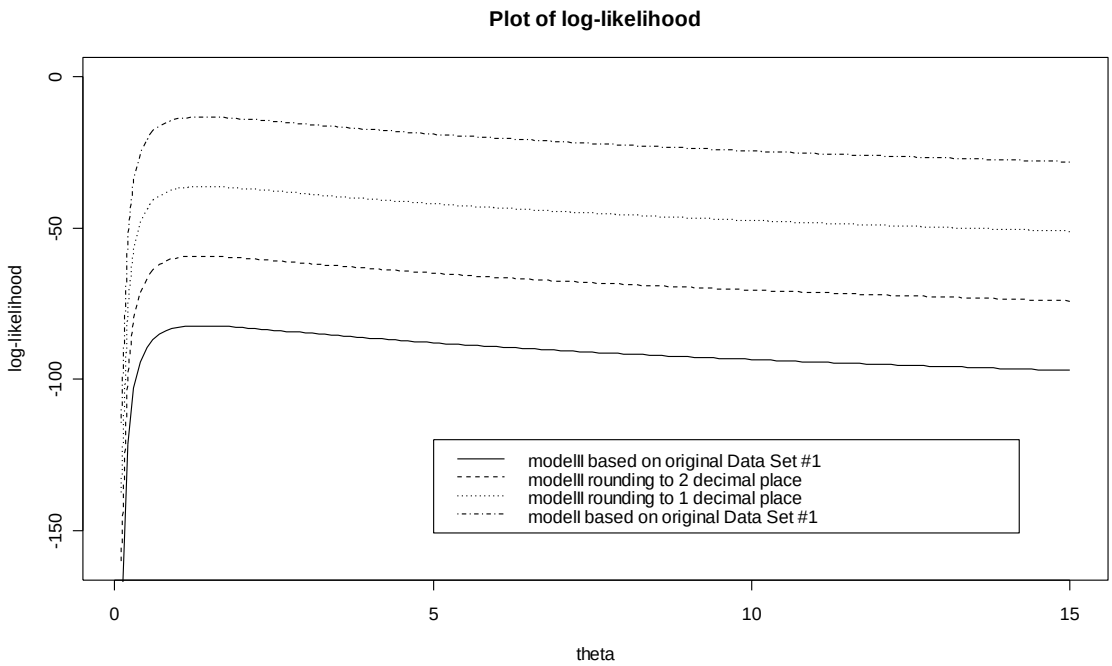
$$L_{.0005}(\theta) = \prod_{i=1}^{10} \left(F(x_i + .0005|\theta) - F(x_i - .0005|\theta) \right)$$

where $F(x|\theta)$ is the exponential *cdf*. Plot the logarithm of $L_{.0005}(\theta)$ for Data Set #1 and compare it to the *loglikelihood* plotted in part (a). Then do something similar, supposing

that Data Set #1 had been recorded and reported to only 2 decimal places, and then to only 1 decimal place. How do these two *loglikelihoods* compare to ln *L.0005 (θ)* and to the one plotted in part (a)? (If this topic of the effect of rounding/digitalization on statistical inference is of interest to you, there are a couple of fairly elementary papers posted on *Vardeman's* home page that you may want to look at. They are the ones to appear in *IEEE Transactions on Instrumentation and Measurement.*)

Denote $L_{o.ooo5}(\theta) = \prod_{i=1}^{10} (F(x_i + 0.0005 \mid \theta) - F(x_i - 0.0005 \mid \theta))$ as model II.

Known that $F(x \mid \theta)$ is the exponential cdf, then $F(x \mid \theta) = 1 - \exp(-x/\theta)$, where $x \geq 0$.



The curves of the log-likelihood function of the model II rounding to 3, 2, and 1 decimal place are all in the same pattern with the one of original model.

2.

(Optional ... not required, but recommended) Suppose that $X_1, X_2, \ldots, X_5$ are iid $N(\mu, \sigma_2)$. Below are two data sets (pseudo-randomly) generated from this model with $\mu = 0$ and $\sigma = 1$.

Data Set #1 : 0.041, −0.705, 0.088, −0.103, −1.9203

Data Set #2 : 0.195, 0.551, −0.821, 0.319, 0.457

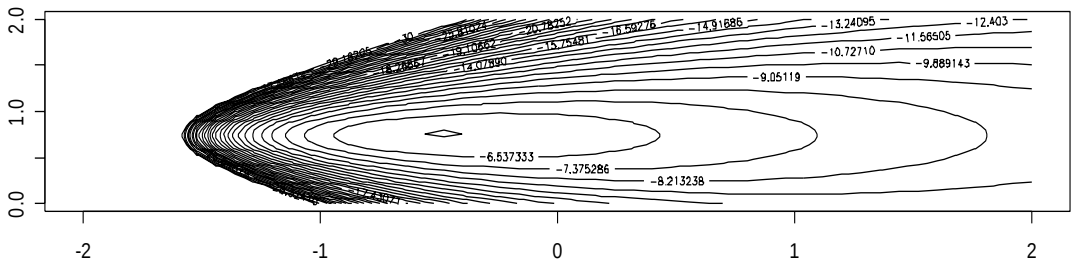
Make contour plots of the *loglikelihood* functions for these two data sets (for $-2 < \mu < 2$ and $0 < \sigma < 2$) (using the definition of likelihood function given in class for jointly continuous observations). Use any software you like to get this done. If you know something about *"R"* you can probably use the hints provided in the *Stat 511* HW #6 assignment posted on the Spring 2004 Stat 511 web page (that you can get to from *Vardeman's* homepage) to figure out the necessary syntax.

Solution:

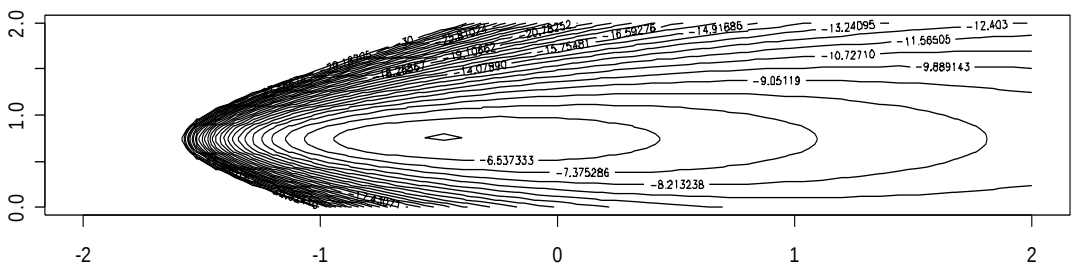
Since $X_1, X_2, ..., X_5$ are $iid\ N(\mu, \sigma^2)$, then the log-likelihood function is

$$\log L(\mu, \sigma^2) = -\sum_{i=1}^{5} \frac{(X_i - \mu)^2}{2\sigma^2} - 5\log(\sqrt{2\pi}\sigma), \text{ where } -2 < \mu < 2 \text{ and } 0 < \sigma < 2.$$

Log-likelihood Contours based on Data Set #1



Log-likelihood Contours based on Data Set #2



3

There are two possible diagnostic tests for a certain medical condition. Individuals from some population (say infants born at hospital *X*) are tested for the condition. In 3 separate studies.

- 10 of 100 tested with Test A alone had a "positive" test result
- 20 of 100 tested with Test B alone had a "positive" test result
- 50 tested with both Test A and Test B broke down as below

		Test B	
		+	−
Test A	+	4	2
	−	8	36

Think of running both tests on an individual as producing an outcome falling naturally into one of the 4 cells of the table below (with corresponding probabilities indicated).

		Test B	
		+	−
Test A	+	p_{++}	p_{+-}
	−	p_{-+}	p_{--}

Since $p_{--} = 1 - (p_{++} + p_{+-} + p_{-+})$, we might think of this as a statistical problem with a 3-dimensional parameter vector $p = (p_{++}, p_{+-}, p_{-+})$. Write an appropriate *loglikelihood* for p based on all data collected in the three studies.

Solution:

In these three studies, the outcome of the first two experiment can by model by $Bin(100, p_{++} + p_{+-})$, and $Bin(100, p_{++} + p_{-+})$ and the outcome of third one can by model by $Multinomial(50, p_{++}, p_{+-}, p_{-+}, p_{--})$.

Since the three studies are mutually independent, the log-likelihood function for these three studies can be expressed as

$$\log L(p \mid x) = \log(\frac{100!}{10!90!}(p_{++} + p_{+-})^{10}(1 - p_{++} - p_{+-})^{90})$$

$$+\log(\frac{100!}{20!80!}(p_{++} + p_{-+})^{20}(1 - p_{++} - p_{-+})^{80}) + \log(\frac{50!}{4!2!8!36!}p_{++}^{4}p_{+-}^{2}p_{-+}^{8}(1 - p_{++} - p_{+-} - p_{-+})^{36})$$

4. (a)

Problem 1.2.2

Consider an experiment in which, for given $\theta = \theta$, the outcome X has density $p(x \mid \theta) = \left(\frac{2x}{\theta^2}\right), 0 < x < \theta$. Let π denote a prior density for θ.

a). Find the posterior density of θ when $\pi(\theta) = 1$, $0 \le \theta \le 1$.

b). Find the posterior density of θ when $\pi(\theta) = 3\theta^2$, $0 \le \theta \le 1$

c). Find $E(\theta \mid x)$ for the two priors in a) and b).

d) Suppose $X_1, \ldots\ldots, X_n$ are independent with the same distribution as X. Find the posterior density of θ given $X_1 = x_1, \ldots\ldots, X_n = x_n$, *when* $\pi(\theta) = 1$, $0 \le \theta \le 1$.

Solution:

a).

$$P(\theta \mid x) = \frac{P(x \mid \theta) P(\theta)}{\int_{\Theta} P(x \mid \theta) P(\theta) d\theta} = \frac{\frac{2x}{\theta^2} \cdot 1}{\int_x^1 \frac{2x}{\theta^2} \cdot 1 d\theta} = \frac{\frac{2x}{\theta^2}}{-\frac{2x}{\theta}\Big|_x^1}$$

$$= \frac{\frac{2x}{\theta^2}}{2-2x} = \frac{x}{\theta^2(1-x)}, \quad x < \theta \le 1$$

b).

$$P(\theta \mid x) = \frac{P(x \mid \theta) P(\theta)}{\int_{\Theta} P(x \mid \theta) P(\theta) d\theta} = \frac{\frac{2x}{\theta^2}(3\theta^2)}{\int_x^1 \frac{2x}{\theta^2}(3\theta^2) d\theta} = \frac{6x}{\int_x^1 6x d\theta} = \frac{1}{1-x}, \; x < \theta \le 1.$$

c). For the prior in a):

$$E(\theta \mid x) = \int_{\theta} \theta P(\theta \mid x) d\theta = \int_x^1 \theta \frac{x}{\theta^2 (1-x)} d\theta$$

$$= \int_x^1 \frac{x}{\theta\ (1-x)} d\theta = \frac{x}{1-x} \ln \theta \Big|_x^1 = -\frac{x}{1-x} \ln x, \quad 0 < x \le 1$$

For the Prior in b):

$$E(\theta \mid x) = \int_{\theta} \theta P(\theta \mid x) d\theta = \int_x^1 \theta \frac{1}{(1-x)} d\theta$$

$$= \frac{\theta^2}{2(1-x)}\Big|_x^1 = \frac{1-x^2}{2(1-x)} = \frac{1+x}{2}, \quad 0 < x \le 1$$

d).

$$P(\theta \mid \underset{\sim}{x}) = -\frac{P(\underset{\sim}{x} \mid \theta) P(\theta)}{\int_{\theta} P(\underset{\sim}{x} \mid \theta) P(\theta) d\theta} \overset{x_1 \ldots x_n \text{ indep.}}{=} \frac{\prod_{i=1}^n P(x_i \mid \theta) \cdot P(\theta)}{\int_{\theta} \prod_{i=1}^n P(x_i \mid \theta) P(\theta) d\theta}$$

$$= \frac{\prod_{i=1}^n \frac{2x_i}{\theta^2}}{\int_{\max_{1 \le i \le n}(x_i)}^1 \prod_{i=1}^n \left(\frac{2x_i}{\theta^2}\right) d\theta} = \frac{\prod_{i=1}^n \frac{x_i}{\theta^{2n}}}{\prod_{i=1}^n x_i \left(\frac{1}{1-2n}\right) \theta^{1-2n} \Big|_{\max_{1 \le i \le n}(x_i)}^1}$$

$$= \frac{1-2n}{\left(1 - \max_{1 \le i \le n} x_i^{1-2n}\right) \theta^{2n}}, \quad \max_{1 \le i \le n} x_i < \theta \le 1$$

Problem 1.2.3

Let X be the number of the failures before the first success in a sequence of Bernoulli trials with probability of success θ. Then $P_\theta[X = k] = (1-\theta)^k \theta, \quad k = 0, 1, 2, \ldots.$ This is

called the *geometric distribution* $G(\theta)$. Suppose that for given $\theta = \theta$, X has the geometric distribution

a). Find the posterior distribution of θ given *X=2* when the prior distribution of θ is uniform on $\{1/4, 1/2, 3/4\}$.

b). Relative to a), what is the most probable value of θ given *X=2* ? Given *X=k* ?

c). Find the posterior distribution of θ given *X=k* when the prior distribution is beta $\beta(r,s)$.

<u>Solutions:</u>

a). Since the prior distribution of θ is uniform on $\{\frac{1}{4},\frac{1}{2},\frac{3}{4}\}$, $\Pi(\theta)=\frac{1}{3}$, if $\theta=\frac{1}{4},\frac{1}{2},\frac{3}{4}$.

Then the posterior density of θ given $X=2$ is

$$P(\theta \mid x=2)=\frac{P(x=2\mid\theta)P(\theta)}{\sum_\theta P(x=2\mid\theta)P(\theta)}=\frac{(1-\theta)^2\theta\cdot\frac{1}{3}}{\left(1-\frac{1}{4}\right)^2\left(\frac{1}{4}\right)\cdot\frac{1}{3}+\left(1-\frac{1}{2}\right)^2\left(\frac{1}{2}\right)\cdot\frac{1}{3}+\left(1-\frac{3}{4}\right)^2\left(\frac{3}{4}\right)\cdot\frac{1}{3}}$$

$$=\frac{(1-\theta)^2\theta}{\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)+\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)}=\frac{16}{5}\theta(1-\theta)^2,\quad \theta=\frac{1}{4},\frac{1}{2}\,or\,\frac{3}{4}$$

b).

$$P(\theta\mid x=2)=\begin{cases}\frac{16}{5}\left(\frac{1}{4}\right)\left(1-\frac{1}{4}\right)^2=\frac{9}{20}, & if\,\theta=\frac{1}{4}\\ \frac{16}{5}\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\right)^2=\frac{2}{5}, & if\,\theta=\frac{1}{2}\\ \frac{16}{5}\left(\frac{3}{4}\right)\left(1-\frac{3}{4}\right)^2=\frac{3}{20}, & if\,\theta=\frac{3}{4}\end{cases}$$

$\theta=\frac{1}{4}$ is the most probable value of θ given *X=2.*

Given *X=k*

$$P(\theta \mid x=k)=\frac{P(x=k\mid\theta)P(\theta)}{\sum_\theta P(x=k\mid\theta)P(\theta)}=\frac{(1-\theta)^k\theta\left(\frac{1}{3}\right)}{\left(1-\frac{1}{4}\right)^k\left(\frac{1}{4}\right)\left(\frac{1}{3}\right)+\left(1-\frac{1}{2}\right)^k\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)+\left(1-\frac{3}{4}\right)^k\left(\frac{3}{4}\right)\left(\frac{1}{3}\right)}$$

$$=\frac{(1-\theta)^k\theta}{\left(\frac{3}{4}\right)^k\left(\frac{1}{4}\right)+\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)+\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)}=\frac{4^{k+1}\theta(1-\theta)^k}{3^k+2^{k+1}+3}$$

$$=\begin{cases}\dfrac{3^k}{3^k+2^{2k+1}+3} & \theta=\dfrac{1}{4}\\[2ex] \dfrac{2^{k+1}}{3^k+2^{k+1}+3} & \theta=\dfrac{1}{2}\\[2ex] \dfrac{3}{3^k+2^{k+1}+3} & \theta=\dfrac{3}{4}\end{cases}$$

If $k=0$, $\theta=\frac{3}{4}$; if $k=1$, $\theta=\frac{1}{2}$; if $k\ge 2$, $\theta=\frac{1}{4}$.

c).

$$P(\theta \mid x=k)=\frac{P(x=k\mid\theta)P(\theta)}{\int_\theta P(x=k\mid\theta)P(\theta)d\theta}=\frac{(1-\theta)^k\theta\frac{1}{\beta(r,s)}\theta^{r-1}(1-\theta)^{s-1}}{\int_0^1(1-\theta)^k\theta\frac{1}{\beta(r,s)}\theta^{r-1}(1-\theta)^{s-1}d\theta}$$

$$=\frac{\theta^r(1-\theta)^{s+k-1}}{\int_0^1\theta^r(1-\theta)^{s+k-1}d\theta}=\frac{1}{\beta(r+1,s+k)}\theta^r(1-\theta)^{s+k-1}\qquad 0\le\theta\le 1$$

$$\theta \mid x=k\sim\beta(r+1,s+k)$$

Problem 1.2.14

In a Bayesian model where $X_1,\ldots\ldots,X_n,X_{n+1}$ *are* i.i.d. $f(x\mid\theta)$, $\theta\sim\pi$, the predictive distribution is the marginal distribution of X_{n+1}, The posterior predictive distribution is the conditional distribution of X_{n+1} given $X_1,\ldots\ldots,X_n$,.

a). If f and π are the $N(\theta,\sigma_0^2)$ and $N(\theta_0,\tau_0^2)$ densities, compute the predictive and posterior predictive distribution.

b). Discuss the behavior of the two predictive distributions as $n\to\infty$.

<u>Solution:</u>

a).

$$f(x_{n+1}\mid\theta)=\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{(x_{n+1}-\theta)^2}{2\sigma_0^2}\right]$$

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\tau_0}\exp\left[-\frac{(\theta-\theta_0)^2}{2\tau_0^2}\right]$$

$$f(x_{n+1}) = \int_{-\infty}^{\infty} f(x_{n+1}\mid\theta)\pi(\theta)d\theta$$

$$= \frac{1}{2\pi\sigma_0\tau_0}\int_{-\infty}^{\infty}\exp\left[-\frac{(x_{n+1}-\theta)^2}{2\sigma_0^2}-\frac{(\theta-\theta_0)^2}{2\tau_0^2}\right]d\theta$$

$$= \frac{1}{2\pi\sigma_0\tau_0}\int_{-\infty}^{\infty}\exp\left[-\frac{\left(\theta-\frac{\tau_0^2x_{n+1}+\sigma_0^2\theta_0}{\sigma_0^2+\tau_0^2}\right)^2+Z^*}{2\frac{\sigma_0^2\tau_0^2}{\sigma_0^2+\tau_0^2}}\right]d\theta$$

$$Z^* = \frac{\tau_0^2x_{n+1}^2+\sigma_0^2\theta_0^2}{\sigma_0^2+\tau_0^2}-\left(\frac{\tau_0^2x_{n+1}+\sigma_0^2\theta_0}{\sigma_0^2+\tau_0^2}\right)^2$$

$$= \frac{\sigma_0^2\tau_0^2(x_{n+1}-\theta_0)^2}{(\sigma_0^2+\tau_0^2)^2}$$

$$f(x_{n+1}) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_0^2+\tau_0^2}}\exp\left[-\frac{(x_{n+1}-\theta_0)^2}{2(\sigma_0^2+\tau_0^2)}\right]$$

So the predictive distribution is,

$$x_{n+1}\sim N\left(\theta_0,\quad \tau_0^2+\sigma_0^2\right)$$

(b)

$$f(x_1,\ldots,x_n,x_{n+1}\mid\theta) = \prod_{i=1}^{n+1}\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{(x_i-\theta)^2}{2\sigma_0^2}\right]$$

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\tau_0}\exp\left[-\frac{(\theta-\theta_0)^2}{2\tau_0^2}\right]$$

$$f(x_1,\ldots,x_n,x_{n+1})$$

$$= \int_{-\infty}^{\infty} f(x_1,\ldots,x_n,x_{n+1}\mid\theta)\pi(\theta)d(\theta)$$

$$= \frac{1}{(2\pi\sigma_0^2)^{\frac{n+1}{2}}\tau_0}\int_{-\infty}^{\infty}\exp[-\frac{\sum_{i=1}^{n+1}(x_i-\theta)^2}{2\sigma_0^2}-\frac{(\theta-\theta_0)^2}{2\tau_0^2}]d\theta$$

$$= \frac{1}{(2\pi\sigma_0^2)^{\frac{n+1}{2}}\tau_0}\int_{-\infty}^{\infty}\exp[-\frac{(\theta-\frac{\tau_0^2\sum_{i=1}^{n+1}x_i+\sigma_0^2\theta_0}{(n+1)\tau_0^2+\sigma_0^2})^2+Z^{**}}{2\frac{\sigma_0^2\tau_0^2}{(n+1)\tau_0^2+\sigma_0^2}}]d\theta$$

$$Z^{**} = \frac{\tau_0^2 \sum_{i=1}^{n+1} x_i^2 + \sigma_0^2 \theta_0^2}{(n+1)\tau_0^2 + \sigma_0^2} - \left(\frac{\tau_0^2 \sum_{i=1}^{n+1} x_i + \sigma_0^2 \theta_0}{(n+1)\tau_0^2 + \sigma_0^2}\right)^2$$

$$= \frac{(n+1)\tau_0^4 \sum_{i=1}^{n+1} (x_i - \bar{x})^2 + \sigma_0^2 \tau_0^2 \sum_{i=1}^{n+1} (x_i - \theta_0)^2}{((n+1)\tau_0^2 + \sigma_0^2)^2}$$

$$f\left(x_1, \ldots, x_n, x_{n+1}\right) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_0^n \sqrt{(n+1)\tau_0^2 + \sigma_0^2}} \exp\left[-\frac{(n+1)\tau_0^4 \sum_{i=1}^{n+1} (x_i - \bar{x}_{n+1})^2 + \sigma_0^2 \tau_0^2 \sum_{i=1}^{n+1} (x_i - \theta_0)^2}{2\left((n+1)\tau_0^2 + \sigma_0^2\right)^2} \right]$$

Similarly,

$$f\left(x_1, \ldots, x_n\right) = \frac{1}{(2\pi)^{\frac{n-1}{2}} \sigma_0^{n-1} \sqrt{n\tau_0^2 + \sigma_0^2}} \exp\left[-\frac{n\tau_0^4 \sum_{i=1}^{n} (x_i - \overline{x_n})^2 + \sigma_0^2 \tau_0^2 \sum_{i=1}^{n} (x_i - \theta_0)^2}{2\left(n\tau_0^2 + \sigma_0^2\right)^2} \right].$$

Then the posterior predictive distribution is,

$$f(x_{n+1} \mid x_1, \ldots, x_n) = \frac{f\left(x_1, \ldots, x_n, x_{n+1}\right)}{f\left(x_1, \ldots, x_n\right)}$$

$$= \frac{\sqrt{n\sigma_0^2 + \tau_0^2}}{\sqrt{2\pi}\sigma_0 \sqrt{(n+1)\sigma_0^2 + \tau_0^2}}$$

$$\exp\{ -\frac{(n+1)\tau_0^4 \sum_{i=1}^{n+1} (x_i - \bar{x}_{n+1})^2 + \sigma_0^2 \tau_0^2 \sum_{i=1}^{n+1} (x_i - \theta_0)^2}{2\left((n+1)\tau_0^2 + \sigma_0^2\right)^2} + \frac{n\tau_0^4 \sum_{i=1}^{n} (x_i - \overline{x_n})^2 + \sigma_0^2 \tau_0^2 \sum_{i=1}^{n} (x_i - \theta_0)^2}{2\left(n\tau_0^2 + \sigma_0^2\right)^2} \}$$

$= \ldots$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_0^2 (1 + \frac{\tau_0^2}{n\tau_0^2 + \sigma_0^2})}} \exp\{ -\frac{(x_{n+1} - \frac{n\tau_0^2 \bar{x}_n + \sigma_0^2 \theta_0}{n\tau_0^2 + \sigma_0^2})^2}{2\sigma_0^2 (1 + \frac{\tau_0^2}{n\tau_0^2 + \sigma_0^2})} \}$$

Therefore, $X_{n+1} \mid X_1, \cdots X_n \sim N(\frac{n\tau_0^2 \bar{x}_n + \sigma_0^2 \theta_0}{n\tau_0^2 + \sigma_0^2}, \sigma_0^2 (1 + \frac{\tau_0^2}{n\tau_0^2 + \sigma_0^2}))$.

<u>Solution II:</u>

Think of writing

$$X_i = \theta + \varepsilon_i$$

, where $\varepsilon_i \underset{\sim}{i.i.d} N(o,\sigma_0^2)$, and $\theta \sim N(\theta_0,\tau_0^2)$ and $\varepsilon_i and \theta$ are independent.

Then

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ X_{n+1} \\ \theta \end{pmatrix} \sim MNV(\theta_0 1, \Sigma)$$

, where $\Sigma = \sigma_0^2 \mathrm{I}_{(n+2)\times(n+2)} + \tau_0^2 J_{(n+2)\times(n+2)}$ and J is a matrix of 1's.

So

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \sim MNV(\theta_0 1, \Sigma_{XX})$$

, where $\Sigma_{XX} = \sigma_0^2 \mathrm{I}_{n\times n} + \tau_0^2 J_{n\times n}$;

Also $X_{n+1} \sim N(\theta_0, \tau_0^2 + \sigma_0^2)$

Let $W = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, and $Z = X_{n+1}$

, then

$W \sim MNV(\theta_0 1, \Sigma_{XX})$, $Z \sim N(\theta_0, \tau_0^2 + \sigma_0^2)$, and $\Sigma_{WZ} = \tau_0^2 1_{(n-1)\times 1}$.

What we want to get is the conditional distribution of $Z \mid W$.

Then by

$$E(Z \mid W) = \mu_Z + \Sigma_{WZ}\Sigma_{WW}^{-1}(W - \mu_W)$$

, and

$$Cov(Z \mid W) = \Sigma_{ZZ} - \Sigma_{ZW}\Sigma_{WW}^{-1}\Sigma_{WZ}$$

, we can get

$$E(X_{n+1} \mid X_1 \cdots X_n) = \frac{n\tau_0^2 \bar{x}_n + \sigma_0^2\theta_0}{n\tau_0^2 + \sigma_0^2}$$

, and

$$Var(X_{n+1} \mid X_1 \cdots X_n) = \sigma_0^2 (1 + \frac{\tau_0^2}{n\tau_0^2 + \sigma_0^2}).$$

Hence, $X_{n+1} \mid X_1, \cdots X_n \sim N(\frac{n\tau_0^2 \bar{x}_n + \sigma_0^2\theta_0}{n\tau_0^2 + \sigma_0^2}, \sigma_0^2(1 + \frac{\tau_0^2}{n\tau_0^2 + \sigma_0^2}))$.

Notice that the inverse of a matrix in the form of

$$aI + bJ$$

has the form of

$$\alpha I + \beta J$$

for $\alpha = \frac{1}{a} and \beta = -\frac{b}{a(a+kb)}$, where k is the dimension of the square matrix I (or J).

Problem 1.3.3

The problem of selecting the better of two treatments or of deciding whether the effect of one treatment is beneficial or not often reduces to the problem of deciding whether $\theta < 0, \theta = 0 or \theta > 0$ for some parameter θ. See Example 1.1.3. Let the actions corresponding to deciding whether $\theta < 0, \theta = 0 or \theta > 0$ be denoted by -1, 0, 1, respectively and suppose the loss function is given by (from Lehmann, 1957)

$\theta \backslash a$	*-1*	*0*	*1*
<0	*0*	*c*	*b+c*
=0	*b*	*0*	*b*
>0	*b+c*	*c*	*0*

, where b and c are positive. Suppose X is a $N(\theta,1)$ sample and consider the decision rule

$$\delta_{r,s}(X) = \begin{cases} -1 & if \quad \bar{X} < r \\ 0 & if\ r < \bar{X} < s \\ 1 & \bar{X} < s \end{cases}$$

a). Show hat the risk function is given by

$$\begin{aligned} R(\theta, \delta_{r,s}) &= c\bar{\Phi}\left(\sqrt{n}(r-\theta)\right) + b\bar{\Phi}\left(\sqrt{n}(s-\theta)\right) && \theta < 0 \\ &= b\bar{\Phi}\left(\sqrt{n}s\right) + b\bar{\Phi}\left(\sqrt{n}r\right) && \theta = 0 \\ &= c\bar{\Phi}\left(\sqrt{n}(s-\theta)\right) + b\bar{\Phi}\left(\sqrt{n}(r-\theta)\right) && \theta = 0 \end{aligned}$$

, where $\bar{\Phi} = 1 - \Phi$, and Φ is the $N(0,1)$ distribution function.

(b) Plot the risk function when b=c=1, n=1 and
(i) *r=-s=-1* and (ii) *r=-1/2s=-1*

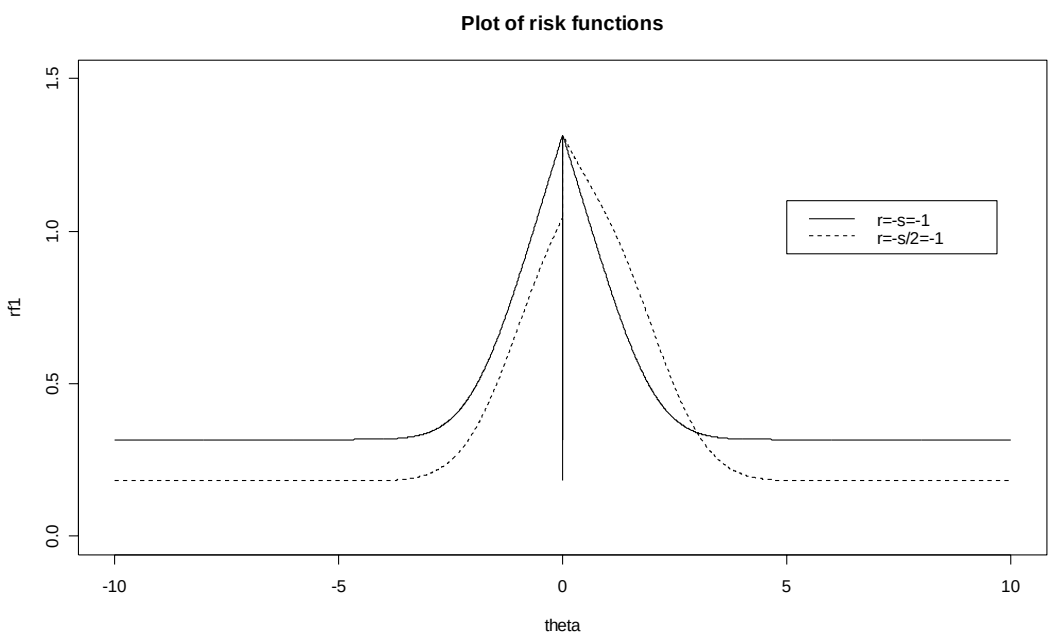
For what values of θ dose the procedure with *r=-s=-1* have smaller risk than the procedure with *r=-1/2s=-1*?

Solution:

$$
\begin{aligned}
V_R(\theta,\delta_{r,s}) &= E_\theta\left(L\left(\theta,\delta_{r,s}(x)\right)\right) \\
&= \begin{cases} 0\cdot P\left(\bar{X}<r\right)+c\cdot P\left(r\le\bar{X}\le s\right)+(b+c)P\left(\bar{X}>s\right) & \theta<0 \\ b\cdot P\left(\bar{X}<r\right)+0\cdot P\left(r\le\bar{X}\le s\right)+(b)P\left(\bar{X}>s\right) & \theta=0 \\ (b+c)\cdot P\left(\bar{X}<r\right)+c\cdot P\left(r\le\bar{X}\le s\right)+0\cdot P\left(\bar{X}>s\right) & \theta>0 \end{cases} \\
&= \begin{cases} c\left[\Phi\left(\sqrt{n}(s-\theta)\right)-\Phi\left(\sqrt{n}(r-\theta)\right)\right]+(b+c)\left[1-\Phi\left(\sqrt{n}(s-\theta)\right)\right] & \theta<0 \\ b\Phi\left(\sqrt{n}(r-\theta)\right)+b\left[1-\Phi\left(\sqrt{n}(s-\theta)\right)\right] & \theta=0 \\ (b+c)\Phi\left(\sqrt{n}(r-\theta)\right)+c\left[\Phi\left(\sqrt{n}(s-\theta)\right)-\Phi\left(\sqrt{n}(r-\theta)\right)\right] & \theta>0 \end{cases} \\
&= \begin{cases} c\left[1-\Phi\left(\sqrt{n}(r-\theta)\right)\right]+b\left[1-\Phi\left(\sqrt{n}(s-\theta)\right)\right] & \theta<0 \\ b\left[1-\Phi\left(\sqrt{n}s\right)\right]+b\Phi\left(\sqrt{n}r\right) & \theta=0 \\ b\Phi\left(\sqrt{n}(r-\theta)\right)+c\Phi\left(\sqrt{n}(s-\theta)\right) & \theta>0 \end{cases} \\
&= \begin{cases} c\bar{\Phi}\left(\sqrt{n}(r-\theta)\right)+b\bar{\Phi}\left(\sqrt{n}(s-\theta)\right) & \theta<0 \\ b\bar{\Phi}\left(\sqrt{n}s\right)+b\Phi\left(\sqrt{n}r\right) & \theta=0 \\ b\Phi\left(\sqrt{n}(r-\theta)\right)+c\Phi\left(\sqrt{n}(s-\theta)\right) & \theta>0 \end{cases}
\end{aligned}
$$

where $\bar{\Phi}=1-\Phi$, and Φ is the $N(0,1)$ distribution function.

(b)



From the plot, we can see that when $\theta \in (0,3)$ the procedure with r=-s=-1 have smaller risk than the procedure with r=-s/2=-1.

Problem 1.3.9

Let θ denote the proportion of people working in a company who have a certain characteristic (e.g., being left-handed). It is known that in the state where the company is located. 10% have the characteristic. A person in charge of ordering equipment needs to estimate θ and uses

$$\hat{\theta} = (.2)(.10) + (.8)\hat{p}$$

, where

$$\hat{p} = \frac{X}{n}$$

is the proportion with the characteristic in a sample of size *n* from the company. Find $MSE(\hat{\theta})$ and $MSE(\hat{p})$. If the true θ is θ_0, for what θ_0 is

$$\frac{MSE(\hat{\theta})}{MSE(\hat{p})} < 1\,?$$

Give the answer for *n*=25 and *n*=100.

<u>Solutions:</u>

$$\hat{\theta} = (.2)(.10) + (.8)\hat{p} \text{ and } \hat{p} = \frac{x}{n}$$

$$X \sim Bin(n, \theta_0),\ E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{EX}{n} = \frac{n\theta_0}{n} = \theta_0$$

$$Var(\hat{p}) = Var\left(\frac{x}{n}\right) = \frac{Var(x)}{n^2} = \frac{n\theta_0(1-\theta_0)}{n^2} = \frac{\theta_0(1-\theta_0)}{n}$$

$$E(\hat{\theta}) = (.2)(.10) + (.8)E(\hat{p}) = 0.02 + 0.8\theta_0$$

$$Var(\hat{\theta}) = 0.64Var(\hat{p}) = 0.64\frac{\theta_0(1-\theta_0)}{n}$$

$$MSE(\hat{p}) = \frac{\theta_0(1-\theta_0)}{n},\quad MSE(\hat{\theta}) = (0.2\theta_0 - 0.02)^2 + 0.64\frac{\theta_0(1-\theta_0)}{n}$$

$$\frac{MSE(\hat{\theta})}{MSE(\hat{p})} < 1 \Rightarrow (0.2\theta_0 - 0.02)^2 + 0.64\frac{\theta_0(1-\theta_0)}{n} < \frac{\theta_0(1-\theta_0)}{n}$$

$$\Rightarrow n\left(0.0004 - 0.008\theta_0 + 0.04\theta_0^2\right) + 0.64\theta_0 - 0.64\theta_0^2 < \theta_0 - \theta_0^2$$

$$\Rightarrow 0.04(n+9)\theta_0^2 - (0.008n + 0.36)\theta_0 + 0.0004n < 0$$

$$\Rightarrow (n+9)\theta_0^2 - (0.2n+9)\theta_0 + 0.01n < 0$$

When *n*=25:

$$34\theta_0^2 - 14\theta_0 + 0.25 < 0$$

$$\Rightarrow 0.0187 < \theta_0 < 0.3931$$

when n=100:

$$109\theta_0^2 - 29\theta_0 + 1 < 0$$
$$\Rightarrow 0.0407 < \theta_0 < 0.2253$$

Problem 1.3.18

For example 1.1.1 consider the loss function (1.3.1) and let δ_k be the decision rule reject the shipment if $X \geq k$.

a). Show that the risk is given by (1.3.7)

b). If *N=10, s=r=1*, $\theta_0 = .1$ and k=3, plot $R(\theta, \delta_k)$ as a function of θ.

c). Same as (b) except k=2. Compare δ_2 and δ_3.

Solution:

(a)

Since the loss function is $\begin{cases} l(\theta,1) = s & if\, \theta < \theta_0 \\ l(\theta,1) = 0 & if\, \theta \geq \theta_0 \\ l(\theta,0) = rN\theta \end{cases}$

Risk function is

$$\begin{aligned} R(\theta,\delta_k) &= E[l(\theta,\delta_k)] \\ &= l(\theta,1)P(\delta_k = 1) + l(\theta,0)P(\delta_k = 0) \\ &= \begin{cases} sP_\theta(X \geq k) + rN\theta P_\theta(X < k) & \theta < \theta_0 \\ rN\theta P_\theta(X < k) & \theta \geq \theta_0 \end{cases} \end{aligned}$$

(b)

For *N=10, s=r=1*, $\theta_0 = .1$ and k=3,

$$R(\theta,\delta_3) = \begin{cases} P_\theta(X \geq 3) + 10\theta P_\theta(X < 3) & \theta < 0.1 \\ 10\theta P_\theta(X < 3) & \theta \geq 0.1 \end{cases}$$

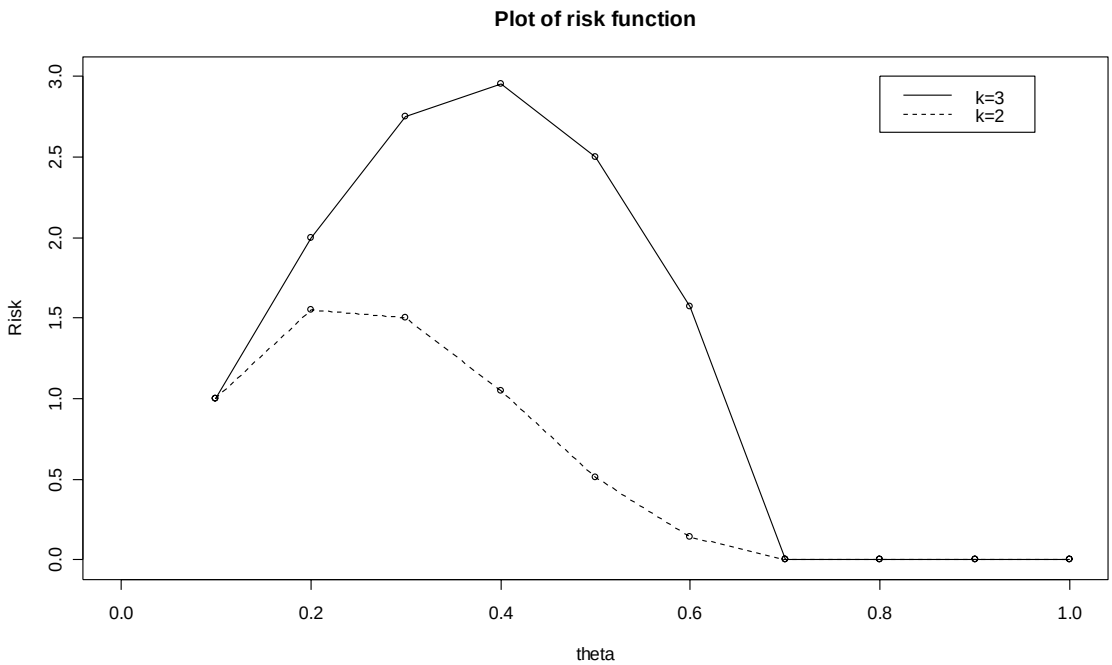
(c)

For *N=10, s=r=1*, $\theta_0 = .1$ and k=2,

$$R(\theta,\delta_2) = \begin{cases} P_\theta(X \geq 2) + 10\theta P_\theta(X < 2) & \theta < 0.1 \\ 10\theta P_\theta(X < 2) & \theta \geq 0.1 \end{cases}$$

Notice in the risk function, $P(X < k) = \sum_{i=0}^{k-1} \frac{\binom{N\theta}{i}\binom{N - N\theta}{n-i}}{\binom{N}{n}}$

, where $\max(n - N(1-\theta), 0) \le i \le \min(N\theta, n)$.



5 (Optional)
Problem 1.2.5

In example 1.2.1 suppose n is large and $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ is not close to 0 or 1 and the prior distribution is beta, $\beta(r,s)$. Justify the following approximation to the posterior distribution

$$P\left[\theta \le t \mid X_1 = x_1, \ldots, X_n = x_n\right] \approx \Phi\left(\frac{t - \tilde{\mu}}{\tilde{\sigma}}\right)$$

where Φ is the standard normal distribution function and

$$\mu = \frac{n}{n+r+s}\bar{x} + \frac{r}{n+r+s}, \quad \tilde{\sigma}^2 = \frac{\tilde{\mu}(1-\tilde{\mu})}{n+r+s}.$$

<u>Solution:</u>

$$f(\theta \mid X) \propto f(X \mid \theta)\pi(\theta)$$

Since $\propto (1-\theta)^{n-n\bar{x}}\theta^{n\bar{x}}(1-\theta)^{s-1}\theta^{r-1}$

$$= \theta^{n\bar{x}+r-1}(1-\theta)^{n-n\bar{x}+s-1}$$

i.e. $\theta \mid X \sim Beta(n\bar{x}+r, n-n\bar{x}+s)$.

Consider the distribution of $\theta \mid \frac{X-\tilde{\mu}}{\tilde{\sigma}}$, where $\tilde{\mu} = \frac{n\bar{x}+r}{n+r+s}$ and $\tilde{\sigma}^2 = \frac{\tilde{\mu}(1-\tilde{\mu})}{n+r+s+1}$.

Denote $\theta \mid X \sim Beta(a,b)$, where $a = n\bar{x}+r$ and $b = n-n\bar{x}+s$.

Known that for $V_1,...,V_a \underset{\sim}{i.i.d} \exp(1)$ and $W_1,...W_b \underset{\sim}{i.i.d} \exp(1)$, $\frac{\sum_1^a V_i}{\sum_1^a V_i + \sum_1^b W_i} \sim Beta(a,b)$.

Then
$$\theta \mid \frac{X-\tilde{\mu}}{\tilde{\sigma}} = \frac{\frac{\sum_1^a V_i}{\sum_1^a V_i + \sum_1^b W_i} - \frac{a}{a+b}}{\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}}$$
$$= \frac{b\sum_1^a V_i - a\sum_1^b W_i}{(a+b)(\sum_1^a V_i + \sum_1^b W_i)\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}}$$
$$= \frac{\frac{b\sum_1^a V_i - a\sum_1^b W_i}{(a+b)}}{\frac{(\sum_1^a V_i + \sum_1^b W_i)}{(a+b)}\sqrt{\frac{ab}{(a+b+1)}}} \qquad (*)$$

Let

$$Y = \frac{b\sum_1^a V_i - a\sum_1^b W_i}{(a+b)}, \text{ and } E(Y) = 0, Var(Y) = \frac{ab}{a+b}.$$

Then (*) equals

$$(*) = \frac{1}{\frac{(\sum_1^a V_i + \sum_1^b W_i)}{(a+b)}} \frac{Y}{\sqrt{\frac{ab}{a+b}}} \frac{\sqrt{\frac{ab}{a+b}}}{\sqrt{\frac{ab}{(a+b+1)}}}.$$

By Central Limit Theorem, $\frac{Y}{\sqrt{\frac{ab}{a+b}}} \to N(0,1)$; and by Weak Law of Large Number,

$\frac{(\sum_1^a V_i + \sum_1^b W_i)}{(a+b)} \to^p 1$, and also $\frac{\sqrt{\frac{ab}{a+b}}}{\sqrt{\frac{ab}{(a+b+1)}}} \to 1$.

So, by Slutskey's Theorem, $\theta \mid \frac{X-\tilde{\mu}}{\tilde{\sigma}} \to N(0,1)$, i.e. $\theta \mid X \to N(\tilde{\mu}, \tilde{\sigma}^2)$.

Thus, $P(\theta \le 0 \mid X_1 = x_1,..., X_n = x_n) \approx \Phi(\frac{t-\tilde{\mu}}{\tilde{\sigma}})$, where Φ is the standard normal distribution function.