MAST30027: Modern Applied Statistics

Week 8 Lab

1. Let $X_1, \cdots, X_n$ be a random sample from a $N(\theta, \sigma^2)$ population, and suppose that the prior distribution on θ is $N(\mu, \tau^2)$. Here we assume that σ^2, μ and τ^2 are all known.

   (a) Find $p(\bar{x}, \theta)$, the joint pdf of $\bar{X}$ and θ.

   Solution: $\bar{X}|\theta \overset{d}{=} N(\theta, \frac{\sigma^2}{n})$. So the joint pdf is

$$p(\bar{x}, \theta) = p(\bar{x}|\theta)p(\theta) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2} \cdot \frac{1}{\tau\sqrt{2\pi}} e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}.$$

   (b) Show that $p(\theta|\bar{x})$ is normal with mean and variance given by $E(\theta|\bar{x}) = \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x} + \frac{\sigma^2}{n\tau^2+\sigma^2}\mu$ and $\text{Var}(\theta|\bar{x}) = \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}$.

   Solution: First one can obtain the following simplification

$$\begin{aligned}
-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2 &= -\frac{n\bar{x}^2}{2\sigma^2} - \frac{1}{2}(\frac{n}{\sigma^2} + \frac{1}{\tau^2})\theta^2 + (\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2})\theta - \frac{\mu^2}{2\tau^2} \\
&= -\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2 + \frac{(n\tau^2\bar{x}+\sigma^2\mu)^2}{2\sigma^2\tau^2(n\tau^2+\sigma^2)} - \frac{n\bar{x}^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} \\
&= -\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2 - \frac{n}{2(n\tau^2+\sigma^2)}(\bar{x}-\mu)^2.
\end{aligned}$$

   Using this and the result in (a), the posterior pdf is

$$\begin{aligned}
p(\theta|\bar{x}) &= \frac{p(\bar{x}|\theta)p(\theta)}{\int_{-\infty}^{\infty} p(\bar{x}|\theta)p(\theta)d\theta} \\
&= \frac{\exp\left\{-\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right\}}{\int_{-\infty}^{\infty}\exp\left\{-\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right\}d\theta} \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n\tau^2+\sigma^2}{\sigma^2\tau^2}}\exp\left\{-\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right\}
\end{aligned}$$

   which is normal with the specified mean and variance. Note that the posterior of θ given $x_1, \cdots, x_n$ is the same as obtained here (we say $\bar{x}$ is sufficient for $\mathbf{x}$, given we know σ^2).

   (c) Show that the marginal pdf of $\bar{X}$, i.e., $p(\bar{x})$, is the pdf of $N(\mu, \frac{\sigma^2}{n} + \tau^2)$.

   Solution: Using (a) and the simplification in (b), it can be found that

$$\begin{aligned}
p(\bar{x}) &= \int_{-\infty}^{\infty} p(\bar{x}|\theta)p(\theta)d\theta \\
&= \int_{-\infty}^{\infty} \frac{\sqrt{n}}{2\pi\sigma\tau}\exp\left\{-\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2 - \frac{n}{2(n\tau^2+\sigma^2)}(\bar{x}-\mu)^2\right\}d\theta \\
&= \frac{1}{\sqrt{2\pi(\tau^2+\sigma^2/n)}}\exp\left\{-\frac{1}{2(\tau^2+\sigma^2/n)}(\bar{x}-\mu)^2\right\}
\end{aligned}$$

   which is the pdf of $N(\mu, \frac{\sigma^2}{n} + \tau^2)$.

   Note: this marginal pdf is not required for finding the posterior pdf of θ.

2. (a) Here is some code for simulating a discrete random variable Y. What is the probability mass function (pmf) of Y?

```
Y.sim <- function() {
  U <- runif(1)
  Y <- 1
  while (U > 1 - 1/(1+Y)) {
    Y <- Y + 1
  }
  return(Y)
}
```

Solution: We have, for $y \geq 1$,

$$\mathbb{P}(Y = y) = 1 - \frac{1}{1+y} - \left(1 - \frac{1}{y}\right) = \frac{1}{y} - \frac{1}{1+y} = \frac{1}{y(1+y)}.$$

(b) Here is some code for simulating a discrete random variable Z. Show that Z has the same pmf as Y.

```
Z.sim <- function() {
  Z <- ceiling(1/runif(1)) - 1
  return(Z)
}
```

Solution:
Put $Z = \lceil 1/U \rceil - 1$ where $U \sim U(0,1)$, then

$$Z = z \iff z < 1/U \leq z + 1 \iff 1/(1+z) \leq U < 1/z.$$

Thus $\mathbb{P}(Z = z) = 1/z - 1/(1+z)$ which is the same as the pmf of Y.

3. Consider the continuous random variable X with pdf given by:

$$f_X(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2} \qquad -\infty < x < \infty.$$

X is said to have a standard logistic distribution. Find the cdf for this random variable. Simulate a sample of size 10 from the distribution using the inversion method. **Solution:** X has cdf

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f_X(y)dy \\ &= \left.\frac{1}{1+e^{-y}}\right|_{-\infty}^{x} \\ &= \frac{1}{1+e^{-x}} \end{aligned}$$

Thus $F^{-1}(y) = \log(y/(1-y)) = -\log((1/y) - 1)$ and we can simulate a sample of size 10 from the distribution of X as follows

```
> set.seed(1000)
> x = -log(1/runif(10) - 1)
> x

 [1] -0.71779506  1.14636578 -2.05114837  0.80365206  0.06563317 -2.62196720
 [7]  1.03929969  0.33730220 -1.29048105 -1.06622107
```