MAST30027: Modern Applied Statistics

Week 7 Lab

1. Suppose that $\theta \sim \text{beta}(a, b)$ and $X \sim \text{bin}(n, \theta)$. That is, θ has pdf

$$f_\theta(x) = \beta(a, b)^{-1} x^{a-1} (1 - x)^{b-1}$$

and $X|\theta = \theta$ has the conditional pmf

$$p_{X|\theta}(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}.$$

(a) Find the joint pdf/pmf of θ and X, and hence show that the condtional pdf of $\theta|X = x$ is beta(u,v), for suitable u and v (depending on x).

(b) Suppose that $a = 2$ and $b = 3$. On the same graph plot the prior distribution of θ, and posterior distributions corresponding to $n = 4$ and $x = 2$, $n = 10$ and $x = 5$, $n = 20$ and $x = 10$, and $n = 50$ and $x = 5$.

To plot beta densities in R you can use `curve`:

```
curve(dbeta(x, 2, 3), 0, 1, ylim=c(0,10))
curve(dbeta(x, 4, 5), 0, 1, col="red", add=TRUE)
```

The beta distribution is called the *conjugate* prior for the binomial, because the posterior is from the same family.

Solution: The joint distribution of X and θ is

$$\begin{aligned} p_{X,\theta}(x, \theta) &= p_{X|\theta}(x) p_\theta(\theta) \\ &= \binom{n}{x} \theta^x (1 - \theta)^{n-x} \beta(a, b)^{-1} \theta^{a-1} (1 - \theta)^{b-1} \end{aligned}$$

The marginal distribution of X is

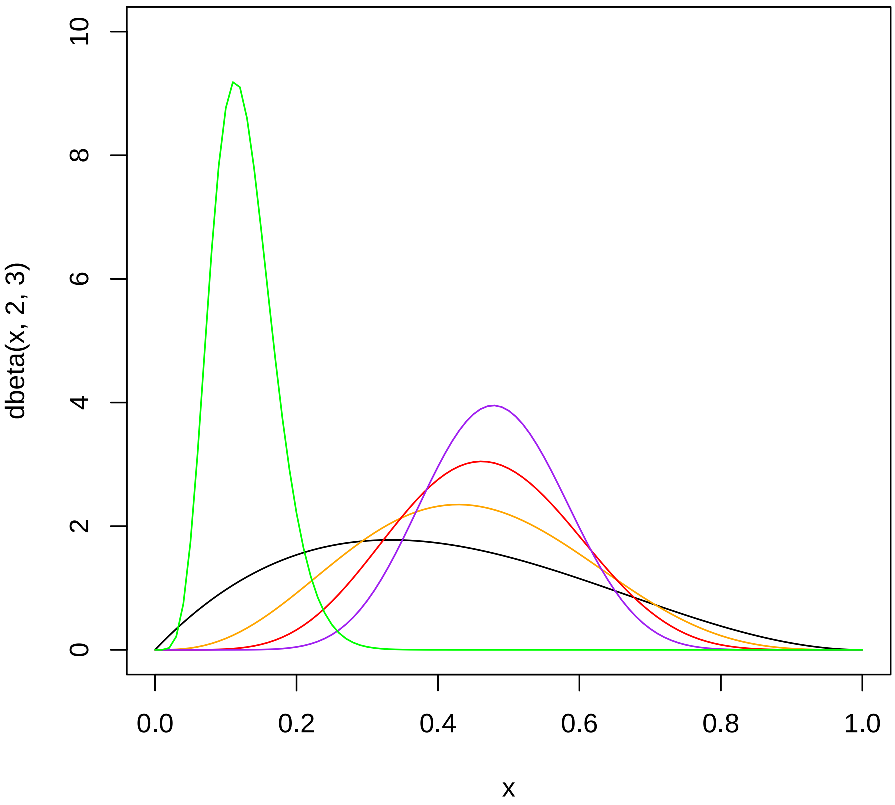
$$p_X(x) = \int p_{X,\theta}(x, \theta) d\theta$$

Thus the conditional distribution of θ given X is

$$\begin{aligned} p_{\theta|X}(\theta) &= \frac{p_{X|\theta}(x) p_\theta(\theta)}{p_X(x)} \\ &\propto p_{X|\theta}(x) p_\theta(\theta) \\ &\propto \theta^{x+a-1} (1 - \theta)^{n-x+b-1} \end{aligned}$$

This is just a beta$(x + a, n - x + b)$ density without its norming constant, so the posterior of θ has a beta$(x + a, n - x + b)$ distribution.

```
> curve(dbeta(x, 2, 3), 0, 1, ylim=c(0,10))  # prior
> curve(dbeta(x, 4, 5), 0, 1, col="orange", add=TRUE)  # posterior n=4, x=2
> curve(dbeta(x, 7, 8), 0, 1, col="red", add=TRUE)  # posterior n=10, x=5
> curve(dbeta(x, 12, 13), 0, 1, col="purple", add=TRUE)  # posterior n=20, x=10
> curve(dbeta(x, 7, 48), 0, 1, col="green", add=TRUE)  # posterior n=50, x=5
```



2. Suppose that θ and X are as above. The marginal distribution of X is given by

$$p_X(x) = \int_0^1 p_{X\theta}(x) f_\theta(\theta) d\theta.$$

X is said to have a beta-binomial distribution.

It is possible, but not easy, to work out p_X for a beta-binomial. However, it is easy to estimate it using simulation.

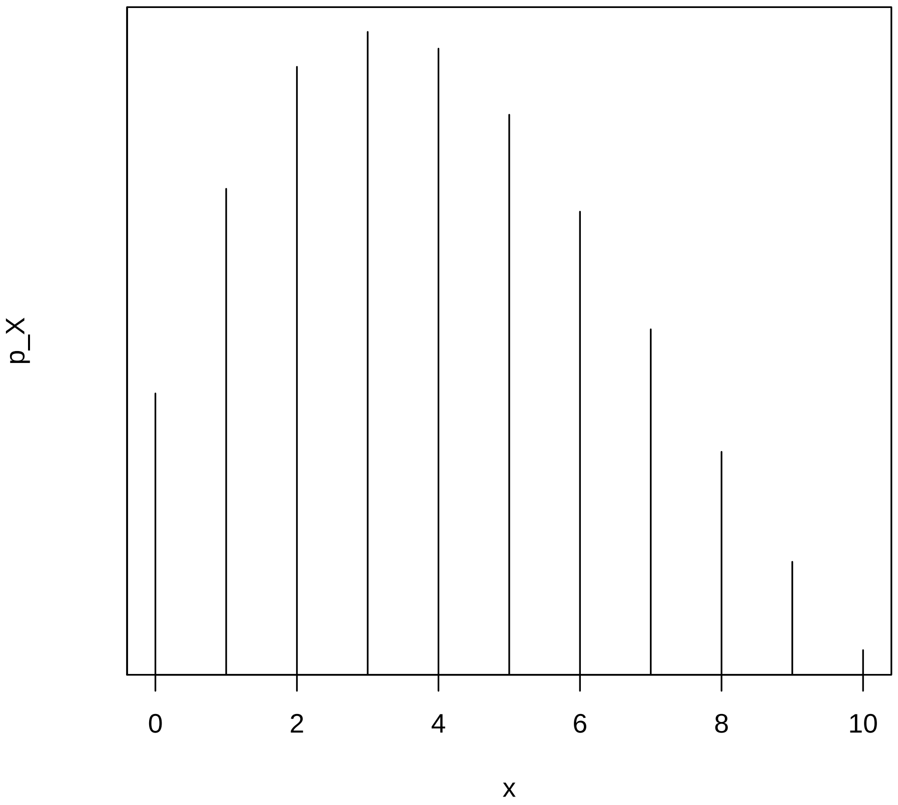
Generate a sample of size 1000,000 from a beta-binomial with $n = 10$, $a = 2$ and $b = 3$. Use it to estimate the pmf of X.

Solution:

```
> N <- 1e6
> tha <- rbeta(N, 2, 3)
> X <- rbinom(N, 10, tha)
> (p_X <- table(X)/N)

X
       0        1        2        3        4        5        6        7
0.066000 0.109845 0.135981 0.143463 0.139880 0.125715 0.104945 0.079740
       8        9       10
0.053489 0.029932 0.011010

> x <- as.numeric(names(p_X))
> plot(x, p_X, type="h")
```



3. 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 the joint pdf of $\bar{X}$ and θ.

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

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

   (b) Show that the posterior pdf of θ, denoted as $p(\theta|\mathbf{x}, \sigma^2, \mu, \tau^2)$, is normal with mean and variance given by $E(\theta|\mathbf{x}) = \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x} + \frac{\sigma^2}{n\tau^2+\sigma^2}\mu$ and $\mathrm{Var}(\theta|\mathbf{x}) = \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}$.

   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

$$p(\theta|\bar{x}, \sigma^2, \mu, \tau^2) \quad = \quad \frac{f(\bar{x}|\theta, \sigma^2)p(\theta|\mu, \tau^2)}{\int_{-\infty}^{\infty} f(\bar{x}|\theta, \sigma^2)p(\theta|\mu, \tau^2)d\theta}$$

$$= \frac{\exp\left\{-\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right\}}{\int_{-\infty}^{\infty}\exp\left\{-\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right\}d\theta}$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n\tau^2+\sigma^2}{\sigma^2\tau^2}}\exp\left\{-\frac{n\tau^2+\sigma^2}{2\sigma^2\tau^2}\left(\theta - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right\}$$

which is normal with the specified mean and variance. Note that the posterior of θ given $x_1, \cdots, x_n$ and (σ^2, μ, τ^2) 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 $\bar{X}$ is $N(\mu, \frac{\sigma^2}{n} + \tau^2)$.
Solution: Using (a) and the simplification in (b), it can be found that

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

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.