Let x be a categorical variable which takes values in $\{0, \ldots, n\}$ and y be a continuous variable such that $y \in (0, 1)$. Let the joint density of (x, y) is

$$f(x,y) = \frac{1}{\text{Beta}(\alpha,\beta)} \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}$$

For $n = 10, \alpha = 2$ and $\beta = 3$, generate samples of (x, y) using two procedures: 1) Full conditional Gibbs and 2) an MH-within-Gibbs procedure, where sample x directly from the full conditional and y using random-walk MH.

We first need $P(x \mid y)$ and $P(y \mid x)$. We calculate these conditionals in two different ways:

Longer and problem-specific route

For this problem, we can get marginally $y \sim$Beta(α, β) as

$P(y = y_1) = \sum_{x=0}^{n} f(x, y_1) = \sum_{x=0}^{n} \frac{1}{\text{Beta}(\alpha,\beta)} \binom{n}{x} y_1^{x+\alpha-1}(1-y_1)^{n-x+\beta-1} = \frac{1}{\text{Beta}(\alpha,\beta)} y_1^{\alpha-1}(1-y_1)^{\beta-1} \sum_{x=0}^{n} \binom{n}{x} y_1^{x}(1-y_1)^{n-x} = \frac{1}{\text{Beta}(\alpha,\beta)} y_1^{\alpha-1}(1-y_1)^{\beta-1}$ as $\sum_{x=0}^{n} \binom{n}{x} y_1^{x}(1-y_1)^{n-x} = 1$ due to binomial density.

Thus $y \sim$Beta(α, β).

Similarly, $x \sim$Bin$(n, \frac{\alpha}{\alpha+\beta})$ as $P(x = x_1) = \int_0^1 f(x_1, y)dy = \int_0^1 \frac{1}{\text{Beta}(\alpha,\beta)} \binom{n}{x} y_1^{x+\alpha-1}(1-y_1)^{n-x+\beta-1} = \binom{n}{x_1} \frac{1}{\text{Beta}(\alpha,\beta)} \int_0^1 y^{x_1+\alpha-1}(1-y)^{n-x_1+\beta-1} = \binom{n}{x_1} \frac{1}{\text{Beta}(\alpha,\beta)} \text{Beta}(\alpha + x_1, \beta + n - x_1)$.

After some simplification of the Beta function in terms of Gamma functions, you get the Binomial density.

Once we have these marginals, we can explicitly get the density expressions of $P(x \mid y) = \frac{f(x,y)}{P(y)}$ and $P(y \mid x) = \frac{f(x,y)}{P(x)}$ as Bin(n, y) and Beta$(x + \alpha, n - x + \beta)$, respectively.

Shorter and general route

To calculate the conditional $P(x \mid y)$, we only pick out the parts of the joint density that involve x (and write it as proportional. The part that involves only y is a constant in this case as we are conditioning on y.) and try to identify if there is a known density that mimics that. Thus, $P(x \mid y) \propto \binom{n}{x} y^x(1-y)^{-x}$ which does not lead to a known density, but if we write $\binom{n}{x} y^x(1-y)^{-x} \propto \binom{n}{x} y^x(1-y)^{n-x}$ (since y is a constant in current case, we can add terms involving only y and still use $\propto$) is a known density Bin(n, y) and thus $x \mid y \sim$Bin(n, y)

To calculate the conditional $P(y \mid x)$, we only pick out the parts of joint density that involve y (and write it as proportional. The part that involves only x is a constant in this case as we are conditioning on x.) and try to identify if there is a known density that mimics that. Thus, $P(y \mid x) \propto y^{x+\alpha-1}(1-y)^{n-x+\beta+1}$ which does not lead to a known density, but if we write $y^{x+\alpha-1}(1-y)^{n-x+\beta+1} \propto \frac{1}{\text{Beta}(x+\alpha,n-x+\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta+1}$ (since x is a constant in current case, we can add terms involving only x and still use $\propto$) is a known density $\text{Beta}(x+\alpha, n-x+\beta)$ and thus $y \mid x \sim \text{Beta}(x+\alpha, n-x+\beta)$

If the proportional expressions, obtained in the above procedure do not lead to any known class of densities even after adding more constant terms (here the word 'constant' is used in a relative sense as in the conditioning variable is also a constant), we can use MH sampling. In the case of MH sampling, we only need to bother about the terms that involve the variable which is being sampled. For example, MH sampling of y from $P(y \mid x)$, we only need $y^{x+\alpha-1}(1-y)^{n-x+\beta+1}$ while calculating the acceptance probability.

In case of transformation of a 'continuous' variable, we need to consider the transformed density while sampling. Like, we transform $y \to \lambda$ such that $y = v(\lambda)$ (in our MH implementation, $v(\lambda) = \frac{e^\lambda}{1+e^\lambda}$). Then we need to accept a move for λ based on the joint density of (x, λ) which is $|\frac{\partial v(\lambda)}{\partial \lambda}| f(x, v(\lambda))$, with the Jacobian adjustment.

However, since x is discrete, any transformation of x does not require a Jacobian adjustment.