1.4 Equation 1.27 gives us the probability density of $p(y)$:

$$p_y(y) = p_x(g(y))|g'(y)|$$

$$\text{where } x = g(y)$$

To find the maximum of the density, we take it's derivative w.r.t y and set it to 0:

$$\frac{dp_y(y)}{dy} = |g'(y)|\frac{dp_x(g(y))}{dy} + p_x(g(y))\frac{d|g'(y)|}{dy}$$

Evaluating the 2 terms separately:

$$|g'(y)|\frac{dp_x(g(y))}{dy} = |g'(y)|\frac{dp_x(g(y))}{dg(y)}\frac{dg(y)}{dy} = |g'(y)|p'_x(g(y))g'(y)$$

$$|g'(y)|\frac{dp_x(g(y))}{dy} = |g'(y)|p'_x(g(y))|g'(y)| = |g'(y)|^2 p'_x(g(y)) \text{ if } g'(y) \geq 0$$

$$|g'(y)|\frac{dp_x(g(y))}{dy} = |g'(y)|p'_x(g(y))(-|g'(y)|) = -|g'(y)|^2 p'_x(g(y)) \text{ if } g'(y) < 0$$

and

$$p_x(g(y))\frac{d|g'(y)|}{dy} = p_x(g(y))g''(y) \text{ if } g'(y) \geq 0$$

$$p_x(g(y))\frac{d|g'(y)|}{dy} = p_x(g(y))(-g''(y)) = -p_x(g(y))g''(y) \text{ if } g'(y) < 0$$

Adding the results for the two terms, we get:

$$\frac{dp_y(y)}{dy} = |g'(y)|^2 p'_x(g(y)) + p_x(g(y))g''(y) \text{ if } g'(y) \geq 0$$

and

$$\frac{dp_y(y)}{dy} = -|g'(y)|^2 p'_x(g(y)) - p_x(g(y))g''(y)$$

$$= -(|g'(y)|^2 p'_x(g(y)) + p_x(g(y))g''(y)) \text{ if } g'(y) < 0$$

The two expressions for the derivative are the same except for the negative sign, which doesn't matter when we set the derivative to 0.

$$\implies |g'(\hat{y})|^2 p'_x(g(\hat{y})) + p_x(g(\hat{y}))g''(\hat{y}) = 0$$

For $x = \hat{x}$, $p(x)$ is maximum, $p'_x(\hat{x}) = 0$. So, for $g(y) = \hat{x}$, the term $|g'(y)|^2 p'_x(g(y))$ goes to 0.

But the term $p_x(g(y))g''(y)$ is not necessarily 0. So, the location $\hat{y}$ of the maximum of the density in y is not in general related to the location $\hat{x}$ of the maximum of the density over x.

In the case of a linear transformation, $g''(\hat{y}) = 0$, so the second term also goes to 0, and the maximum probability for both x and y is related by $\hat{x} = g(\hat{y})$.